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 fourth volume in our series. It collects thirteen working papers by philosophers, political scientists and demographers.

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



Vol. 4

Working paper series 2021:11–23 Editors: Joe Roussos & Paul Bowman

STUDIES ON CLIMATE ETHICS AND FUTURE GENERATIONS

Studies on Climate Ethics and Future Generations Vol. 4

Studies on Climate Ethics and Future Generations Vol. 4

Editors: Joe Roussos & Paul Bowman

Institute for Futures Studies Working Papers 2021:11-23 Stockholm 2021

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.

 \odot The authors and the Institute for Futures Studies 2021

Cover: Matilda Svensson Cover image: Land Salzburg/Franz Wieser Print: Elanders, 2021 Distribution: The Institute for Futures Studies, 2021

Contents

Preface	7
Democratic Representation of Future Generations and Population Ethics Gustaf Arrhenius	11
Global Justice and Future Generations: The Case of Sovereign Wealth Funds Elizabeth Finneron-Burns	27
Children or Migrants as Public Goods? Paul Bou-Habib & Serena Olsaretti	49
Universal Procreation Rights and Future Generations Tim Campbell, Martin Kolk & Julia Mosquera	73
What Is the Point of Offsetting? H. Orri Stefánsson	87
The Role of Subsistence Emissions in Climate Justice Göran Duus-Otterström	101
Climate Change, Corrective Justice, and Non-Human Animals Julia Mosquera	121
Defeating Wrongdoing: Why Victims of Unjust Harm Should Take Priority over Victims of Bad Luck Göran Duus-Otterström & Edward A. Page	143
Review Article: The Ethics of Population Policies Henrik Andersson, Eric Brandstedt & Olle Torpman	167
Optimum Population and Environmental Constraints – A Utilitarian Perspective Partha Dasgupta & S.J. Beard	193
Anonymity and Indefinitely Iterated Addition and Reversal <i>M.A. Roberts</i>	221
Classifying Comparability Problems in a Way that Matters Henrik Andersson & Anders Herlitz	239
Frick's Defence of the Procreation Asymmetry Krister Bykvist & Tim Campbell	263

Preface

Now into its fourth year, the Climate Ethics and Future Generations project is proud to present the fourth volume of its *Climate Ethics and Future Generations* preprint series. The interdisciplinary project, which runs 2018–2023, is led by PI Gustaf Arrhenius and co-PIs Krister Bykvist and Göran Duus-Otterström. The project aims to provide comprehensive and cutting-edge research on ethical questions concerning future generations in the context of climate change policy. It is hosted by the Institute for Futures Studies in Stockholm and is generously financed by Riksbankens Jubileumsfond (the Swedish Foundation for Humanities and Social Sciences). For more information, visit climateethics.se.

The Climate Ethics and Future Generations project has 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. These three themes are duly represented in this volume of thirteen papers, in particular the theme of climate justice.

The volume's opening paper is by PI Gustaf Arrhenius, who explores the intersection of population ethics and democratic theory. Focusing in particular on the issue of the democratic representation of future generations, Arrhenius's main finding is that some versions of the All Affected Principle--the principle that all and only those affected by a democratic decision ought to have a say in it--lead to Total Utilitarianism. Next, Elizabeth Finneron-Burns considers whether it is justifiable for rich states to save resources for its future citizens through so-called sovereign wealth funds, rather than giving these resources to the global poor. Finneron-Burns argues that sovereign wealth funds are not easily reconciled with major theories of distributive justice.

In the volume's third contribution, Paul Bou-Habib and Serena Olsaretti defend the argument that taxpayers should subsidize child-rearing as a public good against the challenge that similar levels of human capital can be attained by increasing skilled immigration. In their contribution, the interdisciplinary group of Tim Campbell, Julia Mosquera (both philosophers), and Martin Kolk (a demographer) argue that intergenerational justice may constrain the moral right for current people to procreate beyond the replacement level, insofar as doing so may limit the opportunities of future people to procreate.

The volume's next four papers focus on questions of corrective justice, particularly as they arise in the context of climate change. H. Orri Stefánsson argues that although individuals have a moral duty to reduce their greenhouse gas emissions, they have no duty to offset their emissions, as opposed to using their money to do good in more effective ways. In his paper, Göran Duus-Otterström investigates the moral status of so-called subsistence emissions, or emissions that are needed to satisfy vital people's interests. While Duus-Otterström agrees with most theorists of climate justice that subsistence emissions are typically morally permissible, he argues, more controversially, that individuals may still incur duties to bear additional mitigation costs in virtue of having produced them.

In her somewhat exploratory paper, Julia Mosquera outlines an account of corrective justice for the harms humans impose on non-human animals, focusing in particular on the case of climate change. Next, Duus-Otterström returns with coauthor Edward A. Page, who argue that there is a reason, based on the impersonal value of corrective justice, to prioritize the alleviation of harms arising from an injustice (like, e.g., many of the harms resulting from climate change) over harms arising from mere bad luck (like those from non-human caused natural disasters).

The volume's next two papers, one by coauthors Henrik Andersson, Eric Brandstedt, and Olle Torpman and the other by coauthors by Partha Dasgupta and S.J. Beard, use major ethical theories to evaluate decision-making that affects the size and structure of populations. While Andersson et. al. review what several major ethical theories imply for population policies, Dasgupta and Beard specifically investigate how utilitarians should approach population policies.

The volume's final three papers are the most theoretical. In her contribution, M.A. Roberts considers a puzzle in population ethics, centering on an anonymity principle for ranking potential populations. Roberts shows how an iterated process of adding a life worth living, and then permuting which lives are at each welfare level, leads to an intuitively repugnant conclusion. She concludes that the unrestricted anonymity principle must be false. Henrik Andersson and Anders Herlitz provide a new way of classifying comparability problems, or cases in which neither of two alternatives is at least as good as the other. Finally, Tim Campbell returns with coauthor and co-PI Krister Bykvist, who in their contribution argue that a recent attempt to account for the so-called Procreation Asymmetry in population ethics-the claim that there is a moral reason not to create a miserable person but no moral reason to create a happy person--is inadequate.

We are pleased to be able to share this new and groundbreaking 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.

> Joe Roussos & Paul Bowman Editors

Gustaf Arrhenius¹

Democratic Representation of Future Generations and Population Ethics²

I shall consider how future people can be represented on the level of ideal theory and what implications such a representation would have for the outcomes of current decisions. More specifically, if future people are represented by proxies, what implications will that have for so-called different number cases, that is, cases where the number of people varies in the compared future populations? I will show a surprising connection between democratic theory and population ethics, and that we can derive principles used in population ethics, such as Total Utilitarianism, from some interpretations of the All Affected Principle, the most plausible solution to the democratic boundary problem.

¹Institute for Futures Studies and Department of Philosophy, Stockholm University; gustaf.arrhenius@iffs.se

² I would like to thank Andrea Asker, Paul Bowman, Krister Bykvist, Tim Campbell, Anders Herlitz, and the audience at the *Democracy and Future Generations Workshop*, IFFS, Stockholm, November 16–17, 2020 for helpful discussions and comments on earlier versions of this paper. Financial support from Riksbankens Jubileumsfond (grant M17-0372:1) and Marianne och Markus Wallenberg Stiftelse (grant MMW 2015.0084) is gratefully acknowledged.

Introduction

I shall consider an extension of justice and democracy that at first might strike one as misguided and bizarre, namely the extension of the demos underlying democratic decision-making to include future generations.

It might strike one as misguided because a theory of democracy is not a theory of justice but rather a practical decision method that is, in turn, justified by an appeal to a theory of justice or some other normative ideal. Here, however, we shall understand democracy as a normative ideal in itself, namely, as an ideal regarding the fair distribution of influence or power (we are not claiming that this is the only plausible way of understanding democracy). As such, it will be a part of a more comprehensive theory of justice.³

This extension might strike one as a bizarre idea since it is impossible for future people to take part in present-day democratic decision-making simply because they are not around. However, that we should include them might follow from the most plausible solution to the democratic boundary problem.⁴ This problem concerns criteria for who should have a right to take part in which decisions in democratic decision making – how to delimit the demos. The most popular and, in my mind, the most promising criterion is the so-called All Affected Principle, which roughly says that the people that are relevantly affected by a decision ought to have, in some sense and to varying degrees, influence over it.

The All Affected Principle has been criticized on several grounds, for example, for proposing what is logically and procedurally impossible or for crowding out individual liberty.⁵ As I have shown elsewhere, these criticisms miss its target.⁶

Another interesting aspect of the All Affected Principle that hasn't been much discussed is that it seems to imply that future generations should have influence over decisions taken today. Since political decisions that we make today will affect the interests of future people, and since some of these effects surely would count as being relevantly affected, the All Affected Principle seems to imply that future generations should have an influence on these decisions, or so the argument goes.

It has been suggested that there are at least two problems with including future generations in current democratic decision-making. The first is that future people are not around so it is impossible to include them in a democratic process. Hence,

³ See Arrhenius (2005), (2018b) for a discussion of both of these understandings of democracy.

⁴ Robert Dahl (1989) refers to this problem as "the problem of the unit" (p. 193), "the problem of inclusion" (p. 119), and sometimes as the "boundary problem" (pp. 146-7). Robert Goodin (2007) calls it "the problem of 'constituting the demos". Frederick G. Whelan calls it "the boundary problem" in his (1983) pioneering article on the subject, and so shall I.

⁵ See e.g., Whelan (1983), Nozick (1974), and Bergström (2005).

⁶ See Arrhenius (2005), (2018b) for a detailed defence.

the All Affected Principle demands what is impossible and should for that reason be rejected as a boundary criterion for democratic decision making.⁷ This "Impossibility Argument" I've discussed elsewhere so we will leave it aside here.⁸ One reason to think that this problem isn't insurmountable is that future people could be represented by proxies, just like present people that for some reason or other cannot directly take part in a decision.⁹

However, not much has been said about a second problem, namely how future people can be represented and what implications such a representation would have for the outcomes of current decisions. This will be the focus of this paper but on the level of ideal theory. More specifically, if future people are represented by proxies, what implications will that have for so-called different number cases, that is, cases where the number of people varies in the compared future populations? These issues are in the domain of ethics known as *population ethics*, which involves foundational questions regarding axiology and our duties to future generations. The main problem in population ethics has been to find an adequate theory about the value or choiceworthiness of outcomes where the number of people, the quality of their lives, and their identities may vary.¹⁰ This paper will show a surprising connection between democratic theory and population ethics, and that we can derive principles used in population ethics, such as Total Utilitarianism, from some interpretations of the All Affected Principle.

The All Affected Principle

The All Affected Principle has been formulated in many different ways, both by its advocates and by those opposing it. To the best of my knowledge, the first formulations of the All Affected Principle are by Robert Dahl and Carl Cohen. The former suggests that "[e]veryone who is affected by the decisions of a government should have the *right to participate* in that government", whereas the latter says that "in a *perfect democracy* all who are thus affected [by a decision] *play some part*". Frederick G. Whelan, in his influential paper on the boundary problem, defines the All Affected Principle as "all those people who are affected by a particular law, policy, or decision *ought to have a voice* in making it".¹¹

⁷ Tännsjö (2007).

⁸ Arrhenius (2018a).

⁹ Kavka & Warren (1983).

¹⁰ The fact that the identities of future people may vary in different outcomes, and that this in many cases is a consequence of our actions, is closely connected to the so-called non-identity problem. See Arrhenius (2000b), (forthcoming); Parfit (1984).

¹¹ Dahl (1970), p. 64; C. Cohen (1971), p. 8; Whelan (1983), p. 16 my emphasis.

The contemporary formulations differ in a somewhat similar manner. In my 2005 and 2018 papers, I formulated the All Affected Principle in terms of influence: "The people that are relevantly affected by a decision ought to have, in some sense and to varying degrees depending on how much they are affected by it, influence over the decision". Likewise for Brighouse & Fleurbaey (although they use the term "power"): "...all individuals with a positive stake should have some power". On the other hand, Ian Shapiro suggests that "[e]veryone affected by the operation of a particular domain of civil society should be presumed to have a say in its governance"; Robert Goodin submits that "all affected interests should have a say"; Lars Bergström claims that "the all-affected principle ... says that every individual who is affected by a given decision should have a vote"; and Torbjörn Tännsjö renders the All Affected Principle as "[e]veryone who is affected by a decision should be allowed to take part in it".¹²

Given such a plethora of formulation, whether the All Affected Principle implies that we should include future generations in present day decision making in any interesting sense depends on which formulation we choose.¹³ We shall thus here focus on two versions that may be interpreted in such a way that they imply that future people should be represented by proxies, just like present people that for some reason or other cannot directly take part in a decision.¹⁴

Cohen formulated the All Affected Principle in terms of "perfect democracy". This suggests a version of the All Affected Principle as part of an ideal-type definition of democratic decision-making. We shall render this idea follows:

The Ideal Democracy Version of the All Affected Principle (AAP-ID): A decision is optimally democratic if and only if each individual's influence on the decision is in due proportion to how each individual's relevant interests are affected by the decision.

Since this condition is devoid of normative content, it doesn't have any implication regarding whether future people *ought* to be included in present decision making.¹⁵ It might only imply that future people must be included for a decision to be optimal-

¹² Arrhenius (2018b), p. 102, cf. Arrhenius (2005); Brighouse & Fleurbaey (2010), p. 139; Shapiro (1996), p. 232; Goodin (2007), p. 50; Bergström (2009), p. 1; Tännsjö (2007), p. 5.

¹³ See Arrhenius (2018a) for a discussion.

 $^{^{14}}$ To the best of my knowledge, Kavka & Warren (1983) was the first to propose this in the democratic theory literature.

¹⁵ It might be partly evaluative depending on how one spell out "interests" (and likewise for

[&]quot;influence"). For example, assume that interests is equated with welfare. To say that Krister has higher welfare than Tim in outcome X is an evaluation to the effect that Krister is *better off* than Tim in X. Still, this wouldn't give the condition any normative bite.

ly democratic. In that case, if we cannot include future people in the required way, then it only follows that no decisions are optimally democratic.

As we said above, one might take democracy as a normative ideal, however, and claim that we ought to approximate it as far as possible:

Normative Ideal Democracy (NID): A decision ought to be as democratic as possible, other things being equal.

The *ceteris paribus* clause leaves open the possibility that our final normative theory might take other important values into account, for example, equality, liberty and welfare.

It is still unclear, however, whether AAP-ID together with NID imply any requirement to the effect that future people ought to be represented in the decision process. It depends on what it is for a decision to be closer to the democratic ideal. However, if we get closer to the ideal by having proxies for those people that are relevantly affected but cannot take part, then AAP-ID together with NID imply that future people should have representatives that take part in present decision-making processes. We shall render this version of the All Affected Principle as follows:

The Representational Ideal Democracy Version of the All Affected Principle (AAP-RID): A decision is optimally democratic if and only if all individuals whose relevant interests are at stake in a decision are represented in the decision procedure by a representative who has influence in proportion to her charge's (represented individual's) stakes.

AAP-RID together with NID imply that future people ought to be represented in the decision process.

Brighouse and Fleurbaey have suggested that "[a]ll individuals should have their interests effectively represented in proportion to their stakes". ¹⁶ Brighouse and Fleurbaey add the "requirement that *whenever possible the individuals should represent themselves*". ¹⁷ When an individual cannot represent herself, then she should be represented by a trustee: "An "appropriate" trustee is one who is the most likely to correctly take account of the ... person's interests." This also holds for future people: "The fact that future generations ... are necessarily out of the decision-making process does not mean that their interests should be neglected."¹⁸

¹⁶ Brighouse & Fleurbaey (2010), p. 150.

¹⁷ Brighouse & Fleurbaey (2010), p. 150.

¹⁸ Brighouse & Fleurbaey (2010), p. 150.

We shall render this version of the All Affected Principle as follows:

The Representational Proportional Version of the All Affected Principle (AAP-RP): All individuals whose relevant interests are at stake in a decision ought to be represented in the decision procedure by a representative who has influence in proportion to her charge's (represented individual's) stakes, other things being equal.

AAP-RP follows from the conjunction of NID and AAP-RID, properly formalized.

Representation of Future Generations

To be able to derive some clear implications from the above principle, we need to say more about how future people are to be represented. Brighouse and Fleurbaey don't say much about this. At some point, they suggest that future people should be represented "by the electorate as a whole" but they unfortunately don't explain why the electorate as a whole would be an "appropriate trustee" for future people's interests.¹⁹

Let me make a suggestion that fits with their idea of an "appropriate trustee" at the level of ideal theory. Each possible future person is represented by a guardian angel who votes in accordance with the interests of that future person. The guardian angel will thus vote for the alternative that is the best one when she is only considering the interests of her charge (the future person under consideration):

Guardian Angel Representation: Each person who cannot adequately exercise her voting right is represented by a guardian angel who votes in accordance with the relevant interests of her charge (the represented person), that is, for the alternative that maximizes the charge's interest satisfaction.

This will include all future people but may also include some present people who cannot adequately exercise their voting right.

In cases involving only the same people in the compared outcomes, it is rather clear how the guardian angel would vote. In cases involving people whose existence is contingent on our choices, however, it is not so clear. An outcome A is better than B for Peter if Peter has higher welfare in A as compared to B. But what if Peter exists in outcome A but not in outcome B? Is it in Peter's interests that outcome A rather

¹⁹ Brighouse & Fleurbaey (2010), p. 150.

than B comes about? Can it be better or worse for a person to be than not to be, that is, can it be better or worse to exist than not to exist at all? This old and challenging existential question has been raised anew in population ethics:²⁰

The Existential Question (in population ethics): Can existence be better or worse for a person than *non*-existence?

Different answers to this question have different implications for the Guardian Angel Representation. The two most discussed answers are:

The Negative Answer: Existence cannot be better or worse than non-existence for a person.²¹

The Affirmative Answer: Existence can be better or worse than non-existence for a person.²²

The Affirmative Answer and an Informal Theorem

Let's first assume the affirmative answer and that a person's relevant interests are understood in terms of a person's welfare (more on the latter below). Then, if the guardian angel has a choice between bringing her charge into existence with negative welfare or not bringing her into existence at all, she would choose the latter. Moreover, if the guardian angel had the choice between bringing her charge into existence with positive welfare or not bringing her into existence, she would choose the former.²³

A democratic theory that represents future people in this manner in combination with AAP-RP, implies, given some further assumption, an interesting result.²⁴ Assume that AAP-RP is satisfied in the following manner:

1. Each person who can adequately exercise a voting right gets a vote with a weight proportional to how her interests are at stake in the decision.

²⁰ See e.g., Arrhenius (2003), (2009), (2015); Arrhenius & Rabinowicz (2010), (2014), (2015); Bykvist (2007), (2015); Bykvist & Campbell (2020); Fleurbaey & Voorhoeve (2015); Holtug (2001), (2010); Johansson (2010); Persson (2017); M. A. Roberts (1998), (2003).

²¹ See e.g. Broome (1999).

²² Arrhenius & Rabinowicz (2010), (2014), (2015) and M. A. Roberts (1998), (2003).

²³ For a detailed exposition of this approach, see Arrhenius & Rabinowicz (2010), (2014), (2015) and Arrhenius (forthcoming).

²⁴ This is an extension of an informal theorem in Brighouse & Fleurbaey (2010) which is restricted to a fixed population setting and doesn't involve representatives for future generations.

- 2. All future persons and all present persons who cannot adequately exercise a voting right are represented by guardian angels who each get a vote with a weight proportional to the stakes of the person she is representing.
- A person's stake (relevant interests) given two alternatives consists in the absolute value of the difference of the (numerical representation of the) person's interests in the two alternatives (measured on an interpersonally comparable ratio scale).²⁵
- 4. If a person exists only in one of the compared alternatives, then her stake is her interests in the alternative in which she exists.
- 5. Every non-represented person votes in accordance with their interests, that is, for the alternative that maximizes their interest satisfaction.
- The weighted majority rule is applied to every pair of alternatives in the decision and that alternatives are ranked according to how many (weighted) votes they get.²⁶

Given the above assumptions, the alternatives are ranked according to the total sum of welfare, that is, along the lines of Total Utilitarianism. Hence, what we could call Welfarist Proportional Representative Democracy (WPRD) will have the same ranking as Total Utilitarianism in population ethics.

Consider for example a choice between two possible future populations, one consisting future people with very high welfare (or interest satisfaction) (population A), and another one consisting of a huge number of people with very low positive welfare (B):

²⁵ Let $I_i(A)$ be a function that returns the numerical representation of individual *i*'s interest satisfaction in outcome A. Then $|I_i(A)-I_i(B)|$ is the weight of individual *i*'s vote for application of the majority rule to the choice between alternative A and B. A ratio scale is unique up to a similarity transformation, which means that the ratios of scale values are preserved. The admissible transformations are all functions of the form $f(x) = \alpha x, \alpha > 0$. Sentences such as "John has many times higher (lower) welfare than Chandra" are meaningful. With such a scale, talk of the total and average amount of welfare in a population makes sense (see F. S. Roberts (1984)).

²⁶ The assumption that everyone votes in her interest can be relaxed: it is sufficient for the theorem that every person votes either in her own interest or for the general good (i.e., for the outcome that maximizes the sum of interest satisfaction).



The blocks in the above diagram represent the two future populations, A and B. The width of each block represents the number of people in the corresponding population; the height represents their lifetime welfare. Dashes indicates that the block in question should intuitively be much wider than shown, that is, the population size is intuitively much larger than shown (in this case population B).

Assume that no person exists in both outcomes and that present people's welfare are not at stake. The guardian angels representing the A-people will vote for A and *vice versa* for the guardian angels representing the B-people. The latter guardian angels will get a much lower voting weight than the ones representing the people in A. However, if B is of a sufficient size, then their guardian angels will, because of their greater number, outvote the guardian angels representing the A-people. Hence, like Total Utilitarianism, Welfarist Proportional Representative Democracy implies Derek Parfit's infamous Repugnant Conclusion.²⁷

The above result holds even if the present people's welfare is at stake. Assume that population A consists of both present and future people, and the same for population B. The present people will vote in favour of A since they have much to lose if B came about. Likewise, the guardian angels representing those future people that exist in both A and B --- the necessary future people --- and those that only exist in A, will vote in favour of A. The guardian angels representing the people that only exist in B, the contingent future people, will get a much lower voting weight than the ones representing the people in A. Again, however, if B is of a sufficient size, then the guardian angels representing the present and necessary future people.

In this sense, it can be said that if we let future people have influence on present decisions, then they are sometimes going to crowd out our influence. This might strike one as the opposite of democracy such that that democracy and the represen-

²⁷ The Repugnant Conclusion is normally formulated and discussed in axiological terms Parfit (1984), Arrhenius (2000a), (forthcoming), but it has also been formulated in terms of choiceworthiness Parfit (1984), Arrhenius (2000b), (2004), (forthcoming).

tation of future people are fundamentally at loggerheads.²⁸ Hence, one can take the above result as a *reductio ad absurdum* of the above idea of representation for future people and Welfarist Proportional Representative Democracy. However, one can also take it as a new reason for accepting the Repugnant Conclusion, derived partly from democratic intuitions.

The Negative Answer

What happens if we go for the negative answer to the existential question? Well, then nothing is at stake for contingent people, that is, people that only exist in one of the two compared outcomes. The guardian angels for contingent people will thus be indifferent between the outcomes since noting is at stake for their charges, and the vote by necessarily existing people will determine the outcome, that is, present people and people who exist in both of the compared outcomes.²⁹ For example, if no one exists both in A and B above (a different people case), then Welfarist Proportional Representative Democracy will now be indifferent between the two populations. Likewise if the present people's welfare is unaffected and all the future people are contingent people.

However, if there is an overlap of people, then A beats B in a vote. Assume that population A consists of both present and future people, and the same for population B. The present people will vote in favour of A since they have much to lose if B came about. The guardian angels representing those future people that exist in both A and B, the necessary future people, will vote in favour of A. The guardian angels representing the people that only exist in A or B, the contingent future people, will be indifferent so we can assume that they will abstain or cast their votes randomly. Hence there will be a clear majority for A over B. And since most future people are contingent relative to some pair of alternatives in a choice situation, it is likely that the present people will rule the world with the negative answer to the existential question.

This of course means that the above theorem yields another conclusion. Instead of yielding Total Utilitarianism it will entail a version according to which alternatives are ranked according to necessary people's total sum of welfare. Hence, we have an interesting derivation of what we could call Necessitarian Total Utilitarianism.

²⁸ See Bergström (2005) for this view.

²⁹ I'm using the terminology of "contingent" and "necessary people" slightly differently as compared to how I define them in Arrhenius (2000b), (2003), (forthcoming) since here they are defined on the two compared populations in a binary choice rather than on the whole choice situation.

Complicating the Picture

In the above discussion, we assumed that the "currency" of the All Affected Principle is people's welfare. This assumption can of course be contested. Notice, however, that the concept of welfare used here can be a broad one. For the present discussion, it doesn't matter whether welfare is understood along the lines of experientialist, desire, or objective list theories.³⁰ Moreover, many of the views presented in the debate on the currency of egalitarian justice as alternatives to welfare, for example Rawls' influential list of primary goods, can be included in the concept of welfare used here.³¹

Still, we need to develop a measure or index of what should count as being relevantly affected by a decision by consulting our considered intuitions about which effects on people's interest are of such significance that they should have a say in a decision, and when some people's interests trump other people's interests.

Such a theory would in many respects be similar to the theories of welfare that have been suggested in the discussion of utilitarianism and to the theories regarding the currency of egalitarian justice. One might also think that one could just import an axiology from this area, such as Rawls' "primary goods" or Sen's "capabilities", as an explication of "relevantly affected". This is suggested by Brighouse and Fleurbaey, and an advantage with this approach is that it might bring democratic decision making more in line with what is good from the perspective of justice and morality.³² However, our judgments about when people are affected by a decision in such a way that they should have influence over it may be different in many respects from our judgments about when people's well-being is affected, or about the relevant goods for the state to distribute in an egalitarian fashion.

The example of "nosy preferences" is a case in point: Even if I am so disgusted by the lewd literature that you read, or by your choice of bedroom activities, that my well-being is seriously at stake, it still seems that I shouldn't have any power over you in regards to such activities. Rather, you (and your partner if one is needed) should have all the power to decide such issues.³³

It might be that the results in the previous section would still follow with the correct currency for the All Affected Principle. It would mainly depend on whether

³⁰ For experientialist theories, see e.g., Sumner (1996), Feldman (1997), (2004), and Tännsjö (1998). For desire theories, see e.g., Barry (1989), Bykvist (1998), Griffin (1986), and Hare (1981). For objective list theories, see e.g., Braybrooke (1987), Hurka (1993), Rawls (1971), and Sen (1980), (1992a), (1993).

³¹ For this debate, see Rawls (1971), Sen (1980), (1992a), (1993), Dworkin (1981b), (1981c), (2000), G. A. Cohen (1989), (1993), Arneson (1989), and Nielsen (1996).

³² Brighouse & Fleurbaey (2010), p. 151. Roughly, if people vote in accordance with what is good for them from the perspective of the metric of social justice, then the winning alternative will also be the one that maximises social justice.

³³ See e.g., Sen (1970); Dworkin (1981a), (2000).

this currency would be measurable on an interpersonally comparable ratio scale. We would then get the similar conclusion but not in terms of welfare but in terms of this currency. So we would get, for example, a version of Total Utilitarianism from the informal theorem above but with a different currency from welfare.

Another complication is that we also have to consider what degree and kind of influence that should be given depending on how one is relevantly affected. This can vary, a point that is often overlooked in the discussion of the All Affected Principle. Sometimes it could be a vote (perhaps with differential weights as we did above), sometimes a veto, sometimes only a right to participate in the deliberation or the right to put forward proposals, sometimes a combination of these and other ways of having influence over a decision. We need to develop a theory regarding what kind of influence or power that should be given to people and guardian angels in different situations, not the least when it comes to existential decisions. So there are indeed further issues to explore regarding the connection between democratic representation of future generations and population ethics.

References

Arneson, R. J. (1989). Equality and Equal Opportunity for Welfare. *Philosophical Studies*, *56*(1), 77–93.

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. (2003). The Person-Affecting Restriction Comparativism and the Moral Status of Potential People. *Ethical Perspectives*, *10*(3–4), 185–195.

Arrhenius, G. (2004). The Paradoxes of Future Generations and Normative Theory. In Jesper. Ryberg & T. Tännsjö (Eds.), *The Repugnant Conclusion: Essays on Population Ethics* (pp. 201–218). Dordrecht: Kluwer Academic Publishers.

Arrhenius, G. (2005). The Boundary Problem in Democratic Theory. In F. Tersman (Ed.), *Democracy Unbound: Basic Explorations* (pp. 71–95). Stockholm: Stockholm University. Filosofiska institutionen.

Arrhenius, G. (2009). Can the Person Affecting Restriction Solve the Problem in Population Ethics? In M. A. Roberts & D. T. Wasserman (Eds.), *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem* (pp. 291–316). Dordrecht; London; New York: Springer.

Arrhenius, G. (2015). The Affirmative Answer to the Existential Question and the Person Affecting Restriction. In I. Hirose & A. Reisner (Eds.), *Weighing and Reasoning: Themes from the Philosophy of John Broome* (pp. 110–125). Oxford ; New York, NY: OUP Oxford.

Arrhenius, G. (2018a). *The All Affected Principle and Future Generations*. Mimeo, Institute for Futures Studies.

Arrhenius, G. (2018b). The Democratic Boundary Problem Reconsidered. *Ethics, Politics & Society: A Journal in Moral and Political Philosophy, 2018*(1), 89–122. https://doi.org/10.21814/eps.1.1.52

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

Arrhenius, G., & Rabinowicz, W. (2010). Better to Be than not to Be? In H. Joas & B. Klein (Eds.), *The Benefit of Broad Horizons: Intellectual and Institutional Preconditions for a Global Social Science* (pp. 399–414). Leiden, Boston: BRILL.

Arrhenius, G., & Rabinowicz, W. (2014). The Value of Existence.

Arrhenius, G., & Rabinowicz, W. (2015). The Value of Existence. In I. Hirose & J. Olson (Eds.), *The Oxford Handbook of Value Theory* (pp. 424–444). Oxford: New York: Oxford University Press.

Barry, B. (1989). Utilitarianism and Preference Change. *Utilitas*, *1*(02), 278–282. https://doi.org/10.1017/S0953820800000273

Bergström, L. (2005). Democracy and Future Generations. In F. Tersman (Ed.), *Democracy Unbound: Basic Explorations I* (pp. 190–192). Stockholm: Filosofiska institutionen, Stockholms universitet.

Bergström, L. (2009). *Principles of Voting Power*. Mimeo, Stockholm University, Department of Philosophy.

Braybrooke, D. (1987). Meeting Needs. Princeton, N.J.: Princeton Univ. Press.

Brighouse, H., & Fleurbaey, M. (2010). Democracy and Proportionality. *Journal of Political Philosophy*, *18*(2), 137–155. https://doi.org/10.1111/j.1467-9760.2008.00316.x

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

Bykvist, K. (1998). *Changing Preferences: A Study in Preferentialism*. Uppsala: Acta Universitatis Upsaliensis.

Bykvist, K. (2007). The Benefits of Coming into Existence. *Philosophical Studies*, *135*(3), 335–362. https://doi.org/10.1007/s11098-005-3982-x

Bykvist, K. (2015). Being and Wellbeing. In I. Hirose & A. E. Reisner (Eds.), *Weighing and Reasoning: Themes from the Philosophy of John Broome* (First edition). Oxford ; New York, NY: Oxford University Press.

Bykvist, K., & Campbell, T. (2020). Persson's Merely Possible Persons. *Utilitas*, *32*(4), 479–487. https://doi.org/10.1017/S0953820820000199

Cohen, C. (1971). Democracy. Athens: University of Georgia Press.

Cohen, G. A. (1989). On the Currency of Egalitarian Justice. Ethics, 99(4), 906–944.

Cohen, G. A. (1993). Equality of What? On Welfare, Goods and Capabilities. In A. Sen & M. C. Nussbaum (Eds.), *The Quality of Life*. Oxford: Clarendon Press.

Cohen, J. (1989). Deliberation and Democratic Legitimacy. In A. P. Hamlin & P. Pettit (Eds.), *The Good Polity : Normative Analysis of the State* (pp. 17–34). Oxford: Basil Blackwell.

Dahl, R. A. (1970). *After the Revolution? Authority in a Good Society*. New Haven: Yale University Press.

Dahl, R. A. (1989). Democracy and Its Critics. New Haven: Yale University Press.

Dworkin, R. (1981a). Is There a Right to Pornography? *Oxford Journal of Legal Studies*, *1*(2), 177–212.

Dworkin, R. (1981b). What is Equality? Part 1: Equality of Welfare. *Philosophy & Public Affairs*, *10*(3), 185–246.

Dworkin, R. (1981c). What is Equality? Part 2: Equality of Resources. *Philosophy & Public Affairs*, *10*(4), 283–345.

Dworkin, R. (2000). *Sovereign Virtue: The Theory and Practice of Equality*. Cambridge, Mass: Harvard University Press.

Dworkin, R. (2002). Sovereign Virtue Revisited. Ethics, 113(1), 106-143.

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

Feldman, F. (2004). *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism*. Oxford: Oxford University Press.

Fleurbaey, M., & Voorhoeve, A. (2015). On the Social and Personal Value of Existence. In I. Hirose & A. Reisner (Eds.), *Weighing and Reasoning* (pp. 95–109). Oxford: Oxford University Press.

Goodin, R. E. (2007). Enfranchising All Affected Interests, and Its Alternatives. *Philosophy and Public Affairs*, *35*(1), 40–68.

Griffin, J. (1986). *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Clarendon Press.

Hare, R. M. (1981). *Moral thinking: Its Levels, Method, and Point*. Oxford: Clarendon.

Holtug, N. (2001). On the Value of Coming Into Existence. *The Journal of Ethics*, *5*(4), 361–384.

Holtug, N. (2010). *Persons, interests, and justice [Elektronisk resurs]*. Retrieved from http://dx.doi.org/10.1093/acprof:oso/9780199580170.001.0001

Hurka, T. (1993). Perfectionism. New York: Oxford University Press.

Johansson, J. (2010). Being and Betterness. Utilitas, 22(03), 285–302.

Kavka, G. S., & Warren, V. (1983). Political Representation for Future Generations. In R. Elliot & A. Gare (Eds.), *Environmental Philosophy: A Collection of Readings*. St. Lucia, Queensland; New York: Pennsylvania State Univ Pr.

Nielsen, K. (1996). Radical Egalitarianism Revisited: On Going Beyond The Difference Principle. *Windsor Yearbook of Access to Justice*, *15*, 121–158.

Nozick, R. (1974). Anarchy, State, and Utopia. Malden, MA: Blackwell.

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

Persson, I. (2017). *Inclusive ethics*. Retrieved from http://dx.doi.org/10.1093/acprof:oso/9780198792178.001.0001

Rawls, J. (1971). A Theory of Justice. Cambridge, Mass.: Belknap Press.

Roberts, F. S. (1984). *Measurement Theory: With Applications to Decision-making, Utility and the Social Sciences*. Retrieved from http://www.gbv.de/dms/bowker/toc/9780521302272.pdf

Roberts, M. A. (1998). *Child versus Childmaker: Future Persons and Present Duties in Ethics and the Law*. Lanham: Rowman & Littlefield Publishers.

Roberts, M. A. (2003). Can it Ever Be Better Never to Have Existed At All? Person-Based Consequentialism and a New Repugnant Conclusion. *Journal of Applied Philosophy*, *20*(2), 159–185. https://doi.org/10.1111/1468-5930.00244

Sen, A. (1970). Collective Choice and Social Welfare. San Francisco: Holden-Day.

Sen, A. (1980). Equality of What? In S. McMurrin (Ed.), *The Tanner Lectures on Human Values*. Cambridge: Cambridge University Press.

Sen, A. (1985). Commodities and Capabilities. Amsterdam: North-Holland.

Sen, A. (1992a). *Inequality Re-examined* (Reprint edition). New York: Harvard University Press.

Sen, A. (1992b). Justice and Capability. In *Inequality reexamined* (pp. 73–101). New York: Russell Sage Foundation; Harvard University Press.

Sen, A. (1993). Capability and Well-Being. In *The Quality of Life*. Oxford : Clarendon Press.

Shapiro, I. (1996). Democracy's Place (1 edition). Ithaca: Cornell University Press.

Sumner, L. W. (1996). Welfare, Happiness, and Ethics. New York: Clarendon Press.

Tännsjö, T. (1998). *Hedonistic Utilitarianism*. Edinburgh: Edinburgh University Press.

Tännsjö, T. (2007). Future People, the All Affected Principle, and the Limits of the Aggregation Model of Democracy. In T. Rønnow-Rasmussen, B. Petersson, J. Josefsson, & D. Egonsson (Eds.), *Hommage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz*. Stockholm University.

Whelan, F. G. (1983). Democratic Theory and the Boundary Problem. In J. R. Pennock & J. W. Chapman (Eds.), *Liberal Democracy* (pp. 13–47). New York: New York University Press.

Elizabeth Finneron-Burns¹

Global Justice & Future Generations: The Case of Sovereign Wealth Funds²

The aim of the paper is to question the justifiability of rich states saving for their future citizens rather than using these resources to alleviate the poverty of the world's current poorest people. Sovereign wealth funds (SWFs), in particular the Norwegian Oil Fund will be used as a case study to bring out the salient issues. I will argue that SWFs like Norway's conflict with the most well-known patterns of distributive justice. After bringing out these conflicts, I examine and reject some potential objections that would nonetheless justify rich states saving for future citizens instead of using the money to alleviate poverty.

¹Western University, efinnero@uwo.ca

² I would like to thank Charles Jones and the audience at the Institute for Futures Studies for helpful discussion and comments. Financial support from Riksbankens Jubileumsfond (the Swedish Foundation for Humanities and Social Sciences) is gratefully acknowledges.

1. Introduction

Eighty percent of the world's population currently lives on less than \$10/day. Meanwhile, rich developed countries are amassing billion- or even trillion-dollar sovereign wealth funds to pass on to their future citizens.

Political theorists have long argued that there are obligations of justice towards the global poor and more recently, they have also advocated that there are duties of justice towards future generations. Although there are well-established literatures in both areas, there is currently little analysis or even recognition of how these two duties may interact. By treating these duties as separate, the intergenerational justice literatures end up treating generations as homogeneous monoliths and ignoring the (vast) inequalities *within* them.

When a state accumulates vastly more wealth than they need, or when their standard of living is extremely high relative to others, global justice scholars argue that justice requires some redistribution.³ At the same time, rich states that save so that their *future* citizens can have very high standards of living are usually lauded as examples of good practice.

One mechanism for saving is to establish a sovereign wealth fund (SWF) into which a state deposits the revenues from the sale their territory's natural resources in order to invest the proceeds and save a portion of the revenue for future members of the state. Although current citizens saving for future members of their state instead of spending the entirety of the resources on themselves seems praiseworthy, I will argue that such state-based saving for the future may not in fact be justified given the current plight of the global poor.

The aim of the paper is to question the justifiability of rich states saving for future citizens rather than using these resources to alleviate the poverty of the world's current poorest people. Sovereign wealth funds, in particular the Norwegian Oil Fund will be used as a case study to bring out the salient issues. Although I will often refer to the Norwegian case, this argument is not really about Norway *per* se. I will argue that Norway's Oil Fund, and indeed any SWF that might be set up by similarly situated countries for similar stipulated purposes, conflicts with the most wellknown patterns of distributive justice. After bringing out these conflicts, I examine and reject some potential objections that would nonetheless justify rich states saving for future citizens instead of using the money to alleviate poverty.

It is worth noting that the arguments in this paper are aimed at those who are already sympathetic to the views that well-off states have obligations of justice to

³ To what extent redistribution is required (e.g. to secure basic subsistence/rights, to promote some currency of global equality, etc.) is a matter of debate, but almost all agree there is at least something owed as a matter of justice.

badly-off states and that there are obligations to future generations. If one rejects either of these premises, then there is no conflict. However, a basic cosmopolitan assumption is that all living human beings, regardless of where they live, are of equal moral concern. I will also assume that all human beings regardless of *when* they live are of equal moral concern. This is a disputed claim, but a plausible one.⁴ People are not owed more or less *merely* because of when or where they live, but they may still be owed more or less for other reasons and this is true intergenerationally as well. Therefore, a distributive principle ought to *prima facie* apply to all people.

2. Sovereign Wealth Funds and the Norwegian 'Oil Fund'

Sovereign wealth funds are state-owned funds that invest the proceeds from a commodity into stocks, bonds, real estate, etc. There are a variety of different purposes for which these funds may be established. My interest here is those funds that are set up explicitly to save for future generations.⁵ Probably the best-known example of such a fund is the Government Pension Fund of Norway, colloquially known as the Oil Fund. Norway established the fund in 1990 to invest the surplus revenues from its oil and gas sector. As of early 2021 it was worth over US\$1.3 trillion, or just under US\$250,000 per current Norwegian citizen. It is the world's largest sovereign wealth fund and owns 1.4 % of all the stocks and shares in the world. The money in the fund comes from taxes levied on petroleum companies operating in Norway, sales of oil exploration licenses, and dividends from the state-owned energy company Equinor.⁶

The Fund's website states that the fund "is saving for future generations in Norway. One day the oil will run out, but the return on the fund will continue to benefit the Norwegian population."⁷ The fund's rules state that a maximum of 3 % of the fund's value can be withdrawn each year,⁸ but in fact the first withdrawals did

⁴ This claim is disputed by the existence of a body of literature debating the merits of a social discount rate—that is, to what extent should future generations' interests be discounted by us today. However, as will become evident later on, accepting that future people's claims are weaker than current people's actually supports the conclusions of the paper. Assuming equality of interests is the harder case.

⁵ Other reasons for establishing SWFs include paying citizens an annual dividend (e.g. Alaska Permanent Fund), investing in higher quality public services (e.g. Texas' Permanent School Fund), or avoiding government revenue or currency volatility (e.g. Hong Kong's Exchange Fund).

⁶ Norges Bank Investment Management (2021), "Government Pension Fund Global,"

www.nbim.no/en/the-fund/ (accessed May 26, 2021).

⁷ Ibid.

 $^{^8}$ This represents a decaying exponential and means it would take at least a generation (23 years) to halve the fund, and in 70 years 10 % of the fund would still remain.

not occur until 2016 and 2017, and by the start of 2018 Norway had returned to making net deposits into the fund. In addition to the deposits, the fund has enjoyed an average annual return of approximately 6 % since its inception.⁹ All this suggests that the fund will continue to grow and be set aside for the use of future generations of Norwegians rather than being used to benefit the people of Norway today.

Sovereign wealth funds such as the Oil Fund are interesting because they represent an altruistic distribution. The Norwegian Bank's investment management arm describes the aim of the Fund as follows: "when oil and gas are no longer available, the fund should have enough reserves that future generations are able to maintain the current high living standard."¹⁰ This short quote gives significant information about the Fund's purpose. It is:

- (i) To be spent on future generations;
- (ii) To ensure future Norwegians have at least as good a quality of life as current Norwegians; and,
- (iii) A form of compensation for having used up the oil/gas reserves which implies that current Norwegians see future Norwegians as having a claim on those resources.

Usually in the global justice literature, we are considering the justification of states using their resources to improve their (existing) citizens' standard of living. In this case there is no suggestion that the Oil Fund will be spent to improve the lives of current Norwegians. Rather, the citizens of Norway are forsaking the money they could have used to invest in infrastructure projects, bring down their high cost of living (e.g. through government subsidies), or directly distribute among citizens (for example, through an annual payment to all citizens like Alaska's sovereign wealth fund¹). They are showing a concern for future generations by forgoing these opportunities for themselves in order to ensure future citizens have the ability to enjoy the same high (or higher) standard of living.

Perhaps it is because this selflessness seems consistent with our views regarding obligations of justice towards future generations and therefore praiseworthy that the ethics of these savings funds has received scant attention from political theorists. Where there is such a discussion, the focus tends to be on the ways the funds are invested—i.e. whether they invest and therefore uphold corrupt regimes,

⁹ Norges Bank Investment Management (2021) "Returns" https://www.nbim.no/en/the-fund/returns/ (accessed May 26, 2021).

 $^{^{10}}$ Norges Bank Investment Management, "Government Pension Fund Global," www.nbim.no/en/the-fund/)#

arms and to bacco manufacturers, and so on—rather than their very existence, and how the money is spent. $^{\rm 11}$

3. The Choice to Make

The question at hand is: is it permissible for current Norwegians to amass and earmark the Oil Fund for the future rich (Norwegians), or does justice require that it is redistributed to the current global poor?

There are various ways of defining the 'global poor'. I will use the figure of US\$10/day or US\$3,650/year as it is clearly insufficient to live a flourishing life, and far below the level enjoyed by even the poorest Norwegian citizens, or indeed citizens of any other developed country. As I said at the outset, 80 % of people in the world live at or below this level, and for the purposes of this paper I will refer to them as the 'Current Poor'. At the other end of the spectrum, in 2019 the median annual income in Norway was US\$51,489.¹² This group will be known as the 'Current Rich'.

There has been a general upwards trajectory in economic development since the Industrial Revolution. In the only sixty years between 1950 and 2010 United States real GDP increased by more than 563 % and the standard of living is expected to double between 2007 and 2100.¹³ Therefore, although we cannot know for certain, I will assume that without savings, future Norwegians (the 'Future Rich') will be at least as well-off as current Norwegians. Even if the fund went bust and lost all value, there is no reason to assume that this would make future Norwegians worse off than current Norwegians since Norway's economy is not currently dependent on petroleum revenues or withdrawals from the Fund anyway. With savings, it is almost certain that future Norwegians will be even better off than current Norwegians. No matter the choice that is made (saving or spending), the Future Rich will be better off than the Current Poor (and, by definition, the Future Poor). Finally, I also assume that giving to the Current Poor will, at a minimum, make the Future Poor at

¹¹ See, for example, Angela L. Cummine, (2016) *Citizens' Wealth: Why (and How) Sovereign Funds Should be Managed by the People for the People.* New Haven: Yale University Press; Hans Erik Naess (2019) "Investment Ethics and the global Economy of Sports: The Norwegian Oil Fund, Formula 1 and the

²⁰¹⁴ Russian Grand Prix" *Journal of Business Ethics* 158: p. 535-546; and Gurneeta Vasudeva (2013) "Weaving Together the Normative and Regulative Roles of Government: How the Norwegian Sovereign Wealth Fund's Responsible Conduct is Shaping Firms' Cross-Border Investments" *Organization Science* 24(6): pp. 1662–1682.

¹² World Population Review (2019) "Median Income by Country",

https://worldpopulationreview.com/country-rankings/median-income-by-country (accessed May 26, 2021)

¹³ Robert J. Gordon, "Is Us Economic Growth Over? Faltering Innovation Confronts the Six Headwinds," (Cambridge, MA: National Bureau for Economic Research, 2012).

least as well off as they otherwise would have been. In other words, it may help them or it may not, but it will not make them any worse off.¹⁴

4. Patterns of Distributive Justice

In this section I will argue that despite seeming admirable, rich states saving for the Future Rich conflicts with commonly-held views about the right pattern of distributive justice, including egalitarianism, prioritarianism, sufficientarianism, and some forms of utilitarianism.¹⁵

Egalitarianism

Telic egalitarianism holds that it is in itself bad if some people are worse off than others. There are six potential relationships of inequality:

- (1) Current Poor vs. Future Rich
- (2) Current Rich vs. Future Rich
- (3) Future Rich vs. Future Poor
- (4) Current Poor vs. Current Rich
- (5) Current Poor vs. Future Poor
- (6) Current Rich vs. Future Poor

Of these, I will only seriously consider 1-4. Because standards of living have increased markedly throughout history. It is likely that even the poorest people today have higher standards of living than most people a thousand years ago. We can expect that this trajectory will continue and that at some point in the future, the Future Poor will be better off than most of us are now, even if it is not guaranteed. We cannot know how saving for the Future Rich compared to spending on the current poor will affect relationship 6. The Future Poor may end up better off than the Current Rich simply by the normal passage of time, regardless of how the money is spent. Or, perhaps, spending the money on the Current Poor would improve the lives of the Future Poor significantly so that they end up better off than the Current Rich. Or, maybe it would make no difference at all. We just don't know. Additionally,

¹⁴ I put non-identity issues aside.

 $^{^{15}}$ The arguments in this section cannot hope to include every iteration of each pattern due to space considerations.

with regard to relationship 5, it is not clear that it would make any difference at all to any inequality between them whether the Oil Fund is saved or spent.¹⁶

The table below shows how the decision about Oil Fund might affect the inequalities in relationships 1-4.

	Parties	Saving for Future Rich	Spending on Current Poor
1	Current Poor &	Increases inequality	Decreases inequality
	Future Rich		
2	Current Rich &	Increases inequality ¹⁷	No change ¹⁸
	Future Rich		
3	Future Rich &	Increases inequality	Decreases inequality
	Future Poor		
4	Current Poor &	No change	Decreases inequality
	Current Rich		

Saving for the Future Rich does not change the well-being of any current people, rich or poor, so it does not change the inequality that exists exclusively between current people (relationship 4). On the other hand, saving for the Future Rich does improve the well-being of the Future Rich, so it will increase inequality between them and the Current and Future Poor (relationships 1 and 3) and probably the Current Rich as well (relationship 2).¹⁵

The alternative, spending the Oil Fund's money on alleviating current global poverty would improve the absolute standard of living of the Current Poor, which would reduce the inequalities in relationships 1, 3, 4. Since neither the Current Rich nor Future Rich will be affected, it would not change the level of inequality in relationship 2.¹⁹

A telic egalitarian who cares about equality for its own sake ought to object to saving the Oil Fund for future generations as it either increases various inequalities (1, 2, and 3) or at least does nothing decrease them (4). Furthermore, spending the Oil Fund on global poverty eradication would decrease inequality across almost all

¹⁶ It's likely, but not certain, that giving to the Current Poor would improve the absolute situation of the Future Poor as well, but unclear if/how it would affect the relative levels of the Current and Future Poor.

¹⁷ Again, because of the trajectory of improvements in standards of living, it's impossible to know exactly what the comparison between the Current and Future Rich will be.

¹⁸ However, if we assumed that the Future Rich will be better off than the Current Rich even in the absence of the Oil Fund savings, then failing to save (i.e. spending on the Current Poor) might actually reduce inequality between the Current Rich and Future Rich too.

¹⁹ However, if we assumed that the Future Rich will be better off than the Current Rich even in the absence of the Oil Fund savings, then failing to save (i.e. spending on the Current Poor) might actually reduce inequality between the Current Rich and Future Rich too.

axes. Of course, some egalitarians are only concerned with equality for instrumental reasons. But those people would likely not be concerned with inequality between non-overlapping generations anyway, since the usual instrumental reasons for valuing equality (viz. domination, unequal status, etc.) do not hold in that case.²⁰ However, spending on the Current Poor reduces inequality between contemporaries (3 and 4), but saving either increases or has no effect, so there are still reasons for instrumental egalitarians to prefer spending on the Current Poor and both forms of egalitarians therefore ought to object to this kind of savings for future generations.

Prioritarianism

The basic tenet of prioritarianism is that benefits to people matter more the worse off people are. We should not give equal weight to equal benefits, whoever receives them. Rather, benefits to the worse off should be given more weight. For our purposes, we know that the Current Poor are worse off than the Current Rich, and that the Future Rich will be at least as well off as the Current Rich.²¹

Current Poor < *Current Rich* ≤ *Future Rich*

The worst off group here is clearly the Current Poor. Therefore, benefits to them matter more than benefits to the Future Rich. Saving the Oil Fund for the Future Rich is tantamount to giving priority to the better off and clearly conflicts with the basic prioritarian principle.

Sufficientarianism

Sufficientarianism holds that it is morally good for as many people as possible to enjoy conditions of life that place them above the threshold that marks the minimum required for a sufficiently good quality of life (however that threshold may be defined). The 'headcount' articulation of sufficientarianism means that we ought to transfer resources from better off people to worse off people when such transfers would increase the total number of people who ever achieve sufficiency. There are other versions of sufficientarianism that argue that we ought to transfer resources from better off people to worse off people when such transfers would help the worse

²⁰ Elizabeth Finneron-Burns (working paper) "Does Intergenerational Equality Matter?"

 $^{^{21}}$ I ignore the position of the Future Poor both because distributing to them is not one the options being considered (so their position in the 'worse off' rankings is irrelevant) and because we do not know where they will fit in relative to the three groups being discussed.

off people get as close to the threshold as possible. In either case, the Oil Fund is inconsistent with plausible versions of sufficientarianism.

Defining the threshold is a problem that has plagued sufficientarian theorists. I will not hope to clearly define what a 'sufficiently good life' is here, but for our purposes, it will be enough to specify at least an upper and lower limit for where the threshold might lie. At the lower end, it seems clear that it must be above the US\$10/day defined as the 'Current Poor' earlier in this paper. People on such a low income are still destitute and unable to enjoy even basic goods like adequate nutrition, let alone the other elements of life (e.g. fulfilling leisure activities, sport, etc.) that most would agree make a life flourishing. So unless one wants to define 'enough' as literally the level at which people can remain alive, the threshold must be higher than that of the Current Poor. At the other end, it cannot be as high as that enjoyed by the Current Rich. The average person in a highly developed country has far more than is necessary to live a flourishing life. If we were to locate the threshold at the median Norwegian income—which is one of the highest in the world—sufficientarianism would cease to provide us with any guidance about distributive justice.²²

According to either version of sufficientarianism, the resources in the Oil Fund ought to be distributed to those below the threshold first. So if we assume that the threshold is somewhere between where the Current Poor and Current Rich are, then the Current Poor are below the threshold and the Future Rich are at or above it (depending on whether we assume the Future Rich will be as well off or better off than the Current Rich respectively). Saving the Fund for future Norwegians ends up distributing to those (who will be) at or above the threshold, rather than those below it.

Utilitarianism

Finally, let us consider utilitarianism. This is the view that we ought to choose the option that will produce the greatest good for the greatest number. A first thought might be that while the number of Current Poor is fixed, there is a potentially infinite number of Future Rich and a sufficient number of future Norwegians would outnumber the existing global poor eventually, meaning that saving the Oil Fund for future Norwegians would help more people than if it were spent on alleviating poverty today.

As I mentioned earlier, there are currently 5.6 billion 'Global Poor' (living on less than US\$10/day) and 5.3 million Norwegians. Therefore, assuming Norway continues

²² For one thing, the headcount view would suggest that those with just slightly lower incomes—still extremely rich comparatively—would be entitled to resources over the very poor.
its current rate of population growth, it would take at least 900 years before there were more future Norwegians than Current Poor. It is unlikely that in 900 years the Fund will still exist in its current form (or at all), so it is improbable that we could say that it would benefit more people by being saved than spent.

But even if it did still exist in 900 years, because of diminishing marginal returns, the benefit to each poor person now is likely higher than the benefit to a future rich person. The same amount of money could provide a poor person with a secure home (for example), and a rich Norwegian with an extra television.

Furthermore, if the money were spent on lifting the Current Poor out of poverty, this could have a ripple effect on increasing the utility of their descendants. The positive utility to be gained from helping the Current (and by extension Future) Poor would extend indefinitely into the future, extending the length of time it would take for the alternative (helping future Norwegians) to outweigh it.

5. Objections

So far I have argued that a sovereign wealth fund such as the Oil Fund that is earmarked for future members of a well-off state conflicts with egalitarian, prioritarian, sufficientarian, and utilitarian theories of justice. However, one could argue that even where a policy conflicts with such principles of distributive justice, the policy may nonetheless be justified by other considerations.

Compatriot Partiality

A critic might object here that there is nothing special about SWFs. States all over the world spend their resources on themselves instead of the global poor, even when doing so conflicts with cosmopolitan principles of distributive justice. Saving for future citizens is just another instantiation of the—in their view, justifiable preference to distribute to compatriots rather than poorer outsiders. In this framing, the people of Norway, current and future, together own the Fund as a collective and are all entitled to benefit from its bounty. To justify compatriot partiality, the critic needs to establish, first, that future Norwegians are plausibly part of the same nation/state, and second, that the criteria for compatriot partiality apply in these circumstances. I refer to the 'people of Norway' because there are two ways to think of them: as members of the *nation* of Norway, or as citizens of the *state* of Norway.

According to David Miller, nations are communities that are:

- 1. Constituted by shared beliefs and mutual commitment. They exist when their members recognise one another as compatriots and believe that they all share characteristics of the relevant kind.
- 2. Extended in history. The nation stretches forward into the future as well as from the past.
- 3. Active in character. Nations are communities that do things together, make decisions, achieve results, etc.
- 4. Connected to a particular territory. A nation exists in a particular place in the world and part of the national identity involves that *particular* place.
- 5. Marked off from other communities by a distinct public culture.²³

As noted in his second criterion, Miller considers nations to be extended through time. That is, they are not merely fleeting entities that come into existence when their members are born and end when their members die. Indeed, it would not make sense to conceive them in that way since, in any given nation, members are being born and dying every day, meaning that the nation would not be the same from one day to the next and it would cease to describe anything meaningful. So, they continue throughout time; however, given Miller's first criterion, it seems unlikely that they extend *indefinitely* into the past or into the future. It is doubtful that long dead members of a given nation necessarily share the same beliefs and commitments.²⁴ The world's values and ways of life have changed significantly even over the past 100 years and it is likely that those in the distant future will have very different views about the good than we do today. Furthermore, it's not clear that criterion three is met with respect to future generations. Future Norwegians do not 'do things together' with current Norwegians, nor are they involved in collective decision-making or other activities.

While I think it is arguable that past, present, and future Norwegians should be regarded as co-nationals, given the stated aims of the Oil Fund, it is clear that current Norwegians do see at least *some* future Norwegians as compatriots. However, would they consider themselves part of the same nation that, at various points in history, was part of multiple different states? Similarly, do current Norwegians consider themselves fellow nationals with those who will live in the very distant future and who may have very little substantively in common with them? None-theless, despite some of these worries, for the purposes of the argument, I will accept

²³ David Miller (1997) On Nationality (Oxford: Oxford University Press).

 $^{^{24}}$ For example, as a British person, I do not feel any sense of shared *nationality* with the makers of Stonehenge, despite inhabiting the same island they did.

that, at least recent past and near future people could be considered part of the same nation as current Norwegians.

If present and future Norwegians are fellow nationals, is prioritising conationals over others justified in the case of the Oil Fund? To answer that question, we can consider Miller's claim that nationals are entitled to prioritise their compatriots over others, when the following conditions are met:

- a. The national relationship must be intrinsically valuable;
- b. The special duties must be integral to the relationship;
- c. The relationship must not inherently involve injustice.²⁵

It is not clear to me that the relationship between current and distant future Norwegians is intrinsically valuable. However, I accept that many people would disagree. Avner de-Shalit, for example, sees communities as transgenerational in nature and the continuation of the community is essential to the transcendence of the self.²⁶ In the case of nations, it also seems that condition two would apply. What is a relationship with future members of the nation if not the existence of certain duties? Jeff McMahan writes that one has a duty to "cherish, sustain, and strengthen the nation and its culture; to preserve its physical treasures and institutions as well as the heritage of its values, traditions, and customs; and to pass these on to subsequent generations."²⁷ The passing along of institutions, cultures, values, etc. seems to be the way in which people form relationships with the future. Likewise, the inheritance of old buildings and traditions seems integral to how we relate to and feel part of the same institution-the nation-as those who were part of it in the past. Finally, although certainly not true of all nations around the world, it is doubtful that the Norwegian nation inherently involves injustice. It is not akin to the mafia or gangs which Miller takes to be examples of inherently unjust groups. Therefore, although I have my doubts about the intrinsic value of the nation transgenerationally, I will also accept that the relationship grounds special duties to future members of the nation. Even accepting this, however, there are reasons why shared nationality need not justify saving the Oil Fund for future compatriots.

However, merely having special obligations to a group does not imply that the special obligation is unlimited. Shared nationality might permit/oblige us to priori-

 ²⁵ David Miller (2007) National Responsibility and Global Justice (Oxford University Press), pp. 34-43.
²⁶ Avner de-Shalit (1994) Why Posterity Matters (Routledge). de-Shalit is discussing communities and not nations specifically, but the arguments he makes can apply to both.

²⁷ Jeff McMahan (1997) "The Limits of National Partiality" in Robert Kim & Jeff McMahan (eds.) *The Morality of Nationalism* (Oxford University Press) p. 130.

tise our fellow nationals but not necessarily in order to benefit them as much as possible. It is more plausible that the obligation is to attend to our own nationals' *needs* first before attending to those of others. On Miller's view, our strongest duty is to ensure that co-nationals have basic needs met. After that, there is a (weaker) remedial responsibility to ensure that non-nationals' basic needs are met, even if we are not personally responsible for their plight. Given that Norway's current citizens already do, and their future citizens are expected to have their basic needs (and much more) secured, Norway has already fulfilled the stronger, nationalist duty. However, remedial responsibility to secure the basic rights of non-nationals remains. Since billions of people in the world do not have their basic rights met, Norway continues to have a remedial responsibility to rectify that. As a result, despite compatriot partiality, the Norwegian nation does not have the right to further benefit themselves, including future Norwegians, until they have fully discharged their remedial responsibilities.

Moreover, the sovereign wealth fund is not really a *national* fund in Miller's sense of the term 'nation'. Rather, it is a fund run by the Norwegian *state*. If the majority of current Norwegians were to emigrate to other states or if there were sufficient immigration into the Norwegian state such that it ceased to be a nation or became a different nation, the Oil Fund would remain in the ownership of the Norwegian government. So even if Norwegians *qua* nation had obligations to future nationals and none to non-nationals, it does not follow that they are entitled to use the resources of Norway *qua state* since these two 'institutions' may not be the same.

This brings us to the second way of interpreting 'the people of Norway': as shared members of the *state* of Norway. Since current and future people are/will be cocitizens of the state of Norway, distributive justice applies between them but not to others. Consequently, it is perfectly legitimate for the state of Norway to earmark their oil revenues for future members of their state, rather than share them with others.

One way global justice theorists have justified this kind of co-citizen partiality is through the reciprocity argument.²⁸ Simon Caney describes it as the view that "persons who engage in a system of cooperation acquire special rights to the goods produced by that cooperation and have entitlements to these goods that non-participants lack" and that states constitute such systems of cooperation. Therefore, fellow citizens have special rights and duties because of their membership in the state.²⁹

²⁸ There are other methods of justification including coercion and shared governance. However, I will not discuss these because I find it implausible that future generations are coerced or commonly governed in the relevant way. For examples of these views, see Michael Blake (2002) "Distributive Justice, State Coercion, and Autonomy" *Philosophy &Public Affairs* 30: 257-296 and Thomas Nagel (2005) "The Problem of Global Justice" *Philosophy &Public Affairs* 33: 113-47.

²⁹ Simon Caney (2005) Justice Beyond Borders (Oxford: Oxford University Press), p. 134.

Are future Norwegians in a reciprocal relationship with current Norwegians? Most of the intergenerational justice literature assumes that they are not since although current people can do many things that affect future generations' wellbeing, they cannot make us better or worse off in turn.³⁰ There are, however, those who disagree with this claim.³¹ My own view aligns with those who reject the idea of intergenerational reciprocity. If they are right, and current and future generations are not in a reciprocal relationship, then saving for future members of the state cannot be justified on the grounds of reciprocity.

However, even if they *are*, in the modern world, current Norwegians are certainly also in a reciprocal system of cooperation with people from all states, including the developing world. The modern world is global. All states engage in international trade, almost all participate in global institutions like the United Nations, NATO, OPEC, and so on, even the most insular ones like North Korea. To paraphrase Simon & Garfunkel, no state is an island.³² The effects of a state's financial and political decisions are felt around the world. If current Norwegians are in a reciprocal relationship with both future Norwegians and current non-Norwegians, it is unclear why reciprocity itself would justify earmarking funds for future members of their state not for existing members of developing states with whom they also cooperate. This means that, at most, there is a justification both for saving for the future and for spending on the poor. It seems reasonable, that given there may be reasons for both, considerations like need should act as a tiebreaker.

Right to Bequest

Another objection is that we should not think of the revenues in the Oil Fund as being owned collectively by all Norwegians in perpetuity, but rather as owned only by current Norwegians. Just as individuals have a putative right to bequest their possessions to the people of their choosing, so too do current Norwegians have the right to bequest the Fund to the people of *their* choosing—i.e. future Norwegians. In other words, perhaps this situation is less like the right to show compatriots partiality, and more like the right to bequest. Perhaps we can think of the current

³⁰ See, for example, John Rawls (1999) *A Theory of Justice, Rev.Ed.* (Oxford University Press): p. 254, Edward Page (2006) *Climate Change, Justice, and Future Generations* (Edward Elgar Press), p. 105, Robert Goodin (1985) *Protecting the Vulnerable: A Reanalysis of Our Social Responsibilities* (Chicago University Press), p. 177, Stephen Gardiner (2004) "The Global Warming Tragedy and the Dangerous Illusion of the Kyoto Protocol" *Ethics and International Affairs*, 18 (1), p. 30 and Derek Parfit (1984) *Reasons and Persons* (Oxford University Press), p. 524-5.

³¹ Axel Gosseries (2009) "Three Models of Intergenerational Reciprocity" in *Intergenerational Justice*, ed. Axel Gosseries and Lukas Meyer (Oxford University Press) and Hugh McCormick (2009) "Intergenerational Justice and the Non-Reciprocity Problem" *Political Studies* 57(2).

³² Simon & Garfunkel (1965) "I Am a Rock" *Sounds of Silence* (Columbia Records).

generation of Norwegians as the 'bequestors', and future Norwegians as the 'bequestees'.

It is generally thought that people have, at a minimum, a pro tanto right to bequest at least some of their wealth to people of their choosing.³³ Imagine a wealthy person dies leaving a one million dollar estate. It is their will that this money be given to their also wealthy child. This counterfactually makes everyone who is not their child worse off than if the money were allocated to the public purse. Many of these other people will already be worse off than the child and would benefit more from the money than the person's child would. Nonetheless, most people would accept that the child is entitled to at least a portion of the million dollars because the parent was entitled to bequest the money to the person of their choice, in this case their child.³⁴ Aren't sovereign wealth funds just bequests writ large? There is a tension between the putative right to bequest to the person(s) of your choosing, and a desire for social equality within a state. Even within a state we accept some right of bequest for reasons I will outline below. So, the analogy goes, even within a world of vast inequality, we might accept an entire state's right to bequest. However, I will argue that this is false because the reasons for allowing personal bequests do not apply on a state-wide scale.

There are a number of reasons one might believe in a right to bequest. First, we might see bequests as part of one's personal prerogative to pursue one's own life plans and goals. Although the receipt of the bequest is only completed after your death, your own well-being may be wrapped up in the well-being of your children. Knowing that they will receive your wealth after you die and will be taken care of may make your own life go better while you are still alive. Second, it might also help your life go better if you trust that the wealth you accumulated will be used for projects that you deem important. This is more likely to happen when the wealth is passed along to family members who feel an obligation to follow your wishes than it is when wealth reverts to the public purse and is redistributed to society at large by the government of the day. Finally, bequests might help you form a special relationship with the next generation if you know that your descendants will inherit what you worked to accumulate. This is particularly relevant in the case of tangible

³³ Of course, some argue that the right is not just *pro tanto*. A libertarian like Nozick would argue that assuming the wealth was acquired through just means, justice in transfer protects a person's right to transfer their wealth to whomever they please. However, it's also unlikely that Nozick would support the obligation to redistribute to poorer countries either, so the tension between global and intergenerational justice does not arise in his theory.

³⁴ The existence of inheritance taxes might be thought to show that many do *not* accept the right to bequest/inherit. However, I think their existence shows that we do accept that right. If we did not, we would not have inheritance taxes, because there would be no such thing as inheritance—all assets with monetary value would revert automatically to the state and wills regarding assets would not exist or be enforceable.

objects such as family farms and heirlooms rather than money. Knowing that your heirlooms or farm will continue in the family cements a special relationship with future owners of these objects and promotes a sense of connection with the future.

These reasons are disputed, even on the individual level³⁵ and it is my contention that none of them apply to the Oil Fund either. First, it is unlikely that redistributing the Oil Fund globally would affect a person's concerns that their descendants be taken care of. A developed country like Norway is unlikely to ever fall into such poverty that its citizens were badly off, Oil Fund or no Oil Fund. As mentioned earlier, Norway is one of the most prosperous countries in the world on the back of its other sources of revenue including hydropower, forestry, fish, and minerals.³⁶ Since oil and gas revenues are not currently being used to benefit current Norwegians, it is unclear why them running out would put future Norwegians' standard of living in jeopardy. Furthermore, it is thought that the oil will in fact never run out because of reduced global demand due to concerns about carbon emissions. All this points to a likelihood that Norway will develop new and existing sources of national revenue to provide for its citizens in the long term anyway, even if they were not permitted to pass along the Oil Fund to future generations of Norwegians.

Second, unlike with individual bequests, there is no guarantee that future Norwegians would use the fund the way that current Norwegians intend it. When individuals specify conditions of inheritance in their wills, they are usually enforceable by the courts (with some limited exceptions). On the other hand, with a sovereign wealth fund, although the governmental act creating the fund may specify its future, the Norwegian state is free to change the law at any time. Even constitutions, the highest law of the land, include modification mechanisms. So the argument that people have an interest in ensuring their bequests are used for certain purposes is not analogous to the case of a sovereign wealth fund.

Furthermore, the state money in the Oil Fund owned by the current generation of Norwegians as a collective does not affect the generation's ability to live a flourishing life or have a personal relationship with the next generation since most Norwegians have no personal connection to the money in the fund. It is not something they worked hard for—after all, they just happened to be born inside a territory that lays claim to large oil deposits—and it is not something to which there is an emotional connection. This is not to say that Norwegians would not feel a *loss* were it redistributed elsewhere; they might well feel annoyed or frustrated. But it would not be an emotional loss like that of a precious heirloom or family farm where

³⁵ Daniel Halliday (2018) The Inheritance of Wealth (Oxford University Press).

³⁶ 'Oil and Gas Taxation in Norway' (2014)

https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Energy-and-Resources/gx-er-oil-and-gas-taxguide-norway.pdf_accessed January 22, 2021

generations of family members had grown up and raised their own families in the farming tradition.

Entitlement to Resources

There is a more fundamental response available to both the compatriot partiality and right to bequest justifications. That is, both assume that Norway (understood as either an intergenerational collective or as just the current generation) is entitled to decide what happens with the oil revenues in the first place. As Chris Armstrong notes:

[t]he idea that Norwegians in thirty or fifty years' time ought to enjoy standards of living that Norwegians presently do, and that current generations should not fritter away the windfall of the North Sea gas and oil, sounds like an admirable one. However, it would presumably be considerably less admirable if this project of intergenerational justice is bankrolled by using resource wealth to which Norwegians have *weak or non-existent claims* [my emphasis].³⁷

Armstrong argues that these funds are derived from selling natural resources and that the distribution of natural resources itself is often seen as a question of global justice. It is not, in other words, to be taken for granted that Norway has an automatic right to the proceeds from the natural resources that happen to be within the territory it controls, and as a result, it should not be assumed that they have a subsequent entitlement to enjoy the full value of the Oil Fund either now or in the future. So the first step in justifying Norway's right to allocate the funds in the SWF must be to establish whether Norway is entitled to the profits from the oil in the first place. If they are not, then they are likely not entitled to the money in the Oil Fund either.

Armstrong argues that Norwegians do not, in fact, have an exclusive claim over the oil or to the money in the Oil Fund. He suggests there are three reasons we might see a state having a special claim over their resources:

- (1) *Improvement*. They have acted to increase the value of the resource and are therefore entitled to the increased value (the difference between what it was worth before and after the improvement) only.
- (2) *Attachment*. They have formed life plans or projects that depend up on continuing access to the natural resource.

³⁷ Chris Armstrong, "Sovereign Wealth Funds and Global Justice," *Ethics & International Affairs* 27, no. 4 (2013): 416.

(3) *Self-Determination*. Without control over the resource the state would no longer be able to exercise effective self-determination.

Armstrong goes on to argue persuasively that none of these three reasons hold in the case of oil or the proceeds from oil. In terms of *Improvement*, since the revenue comes, not from improving crude oil, but from selling others the rights to do so, Norway itself has not improved the resource and is not entitled to its increased value. With regards to Attachment, Armstrong concedes that Norwegians may feel attached to the oil under their territory, but that it is not *oil* that is in the fund, "it is composed of the proceeds of *selling* petroleum to the highest bidder. If that petroleum was so non-substitutably important to Norwegians, why would they sell it in the first place?"³⁸ Although Norwegians may have structured their plans around the money in the fund, Armstrong argues, money is not the kind of thing to which someone can become attached in the necessary sense to generate a special claim. Finally, he argues that Self-Determination is also not able to justify Norway's special entitlement to the funds. The reason given for Self-Determination is that having control over a state's natural resources is necessary in order to secure its citizens' basic rights. However, as Armstrong points out, this reason does not generate a special claim once basic rights have been secured-in other words, it does not justify states controlling any surplus natural resources. Certainly, in the case of Norway, its citizens' basic rights are more than minimally secured.

As a result, there seems to be strong reasons to deny Norway's entitlement to the revenues in the Fund at all. If this is right, then neither compatriot partiality nor the right to bequest can be justifications for Norway saving for its own future citizens. After all, one cannot permissibly show partiality with or have a right to give what one does not rightfully hold.

The Difference Principle

A final way a critic might justify savings for future citizens through SWFs is to argue that wealthy states/nations are permitted to establish these SWFs and save so long as they invest the funds in developing countries. Although in reality the Fund explicitly avoids investing in equities, real estate, or government bonds in developing countries³⁹, if the Oil Fund were to do the opposite and invest in companies/governments in developing countries, this would, so the argument goes, help both parties.

³⁸ Chris Armstrong, "Sovereign Wealth Funds and Global Justice"

³⁹ Norges Bank Investment Management, "How We Invest", https://www.nbim.no/en/the-fund/how-we-invest/ (accessed May 20, 2021)

Although inequality would remain between the rich and poor, the poor would, arguably, be better off than if the SWF were not allowed at all.

There are a few problems with this argument. First, remember G.A. Cohen's powerful criticism of the difference principle itself. Norway's options are not just 1) establish the Oil Fund and invest in developing countries, keeping profits for the people of Norway (as proposed by the critic); or 2) do not establish the Oil Fund at all and spend proceeds on current Norwegians. There is, of course, a third option: establish the Fund and spend the revenues on alleviating global poverty. The difference principle argument assumes that rich countries like Norway require the incentives in order to improve the lot of the worse off.

Second, there is a practical concern that in most countries, the companies that are listed on stock exchanges are large, often multinational corporations. While the argument may appeal to those who accept the theory of 'trickle down economics', the reality is that investing in such large corporations may have no positive impact whatsoever on anyone other than shareholders and the highest levels of management who are, of course, already very well off. Rather than creating jobs and lifting people out of poverty, the extra investment may simply line the pockets of the rich.

There are also questions about how this version of the difference principle would be derived. Although there is some debate about whether he should have or not, Rawls explicitly rejected a global difference principle. What is this proposal if not that? However, we may think that this proposal is substantively different from a global difference principle, or that Rawls was wrong in rejecting a global difference principle. Even so, it would have to be derived from some form of original position, and as has been shown elsewhere, there is no composition of the original position that can adequately cope with intergenerational issues.⁴⁰ In brief, if you imagine that the original position is composed of people from a single generation, there is no reason for them to save—what's in it for them? If you include people from a variety of generations and blind them from knowing when their generation will exist, the spectre of the non-identity problem arises.

6. Conclusions

I have argued that there is a tension between our duties of justice to future generations and the members of the global poor that has heretofore remained unremarked on. That is, by discharging putative duties to save for future generations, we may be

⁴⁰ Elizabeth Finneron-Burns (2016) "The Intergenerational Original Position" *Social Theory & Practice* 43(4) and Gustaf Arrhenius and Emil Andersson (2021) "Constructivist Contractualism and Future Generations" in Stephen Gardiner (ed.) *Oxford Handbook of Intergenerational Ethics* (Oxford University Press).

forsaking our duties to alleviate the poverty of the global poor. This tension was explored through the lens of national sovereign wealth funds, and in particular the Norwegian Oil Fund.

Although sovereign wealth funds seem like laudable ideas since they fulfil our putative duties to future generations, they conflict with cosmopolitan iterations of egalitarian, prioritarian, sufficientarian, and utilitarian theories of justice—that is, when we consider that every person should be counted equally regardless of when in time or where in the world they live. I then considered and rejected various justifications for partiality nonetheless.

In the specific case of Norway, my analysis suggests that they ought to spend the revenues in the Oil Fund on discharging their duties to the existing global poor. How these duties are filled in is a matter that is already under debate in the literature—they could be egalitarian, sufficientarian, etc. Although my personal inclination tends towards a sufficientarian view, the duties may indeed be much stronger.

References

Alaska Permanent Fund Corporation, https://apfc.org/ (accessed May 6, 2021).

Armstrong, Chris. "Sovereign Wealth Funds and Global Justice". *Ethics & International Affairs* 27, no. 4 (2013): 413-28.

Arrhenius, Gustaf and Emil Andersson (forthcoming) "Constructivist Contractualism and Future Generations" in Stephen Gardiner (ed.) *Oxford Handbook of Intergenerational Ethics* (Oxford University Press).

Blake, Michael (2002) "Distributive Justice, State Coercion, and Autonomy" *Philosophy & Public Affairs* 30.

Caney, Simon (2005) Justice Beyond Borders (Oxford: Oxford University Press)

Cummine, Angela (2016) *Citizens' Wealth: Why (and How) Sovereign Funds Should be Managed by the People for the People* (Yale University Press)

Deloitte Touche Tohmatsu (2014), 'Oil and Gas Taxation in Norway' https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Energyand-Resources/gx-er-oil-and-gas-taxguide-norway.pdf (accessed January 22, 2021)

de-Shalit, Avner (1994) Why Posterity Matters (Routledge)

Finance, Norwegian Ministry of. "The Government Pension Fund 2018: Report to the Storting". 2018.

Finneron-Burns, Elizabeth (2016) "The Intergenerational Original Position" Social Theory & Practice

Finneron-Burns, Elizabeth (working paper) "Does Intergenerational Equality Matter?"

Gardiner, Stephen (2004) "The Global Warming Tragedy and the Dangerous Illusion of the Kyoto Protocol" *Ethics and International Affairs*, 18 (1)

Global Issues, "Poverty Facts and Statistics" http://www.globalissues.org/article/26/poverty-facts-and-stats (accessed May 26, 2021).

Goodin, Robert (1985) *Protecting the Vulnerable: A Reanalysis of Our Social Responsibilities* (Chicago University Press)

Gordon, Robert J. (2012) "Is US Economic Growth Over? Faltering Innovation Confronts the Six Headwinds," (Cambridge, MA: National Bureau for Economic Research).

Gosseries, Axel (2009) "Three Models of Intergenerational Reciprocity" in *Intergenerational Justice*, ed. Axel Gosseries and Lukas Meyer (Oxford University Press)

Halliday, Daniel, "Inheritance of Wealth: Justice, Equality, and the Right to Bequeath (OUP) 2018.

McCormick, Hugh (2009) "Intergenerational Justice and the Non-Reciprocity Problem", *Political Studies* 57(2).

McMahan, Jeff (1997) "The Limits of National Partiality" in Robert Kim & Jeff McMahan (eds.) *The Morality of Nationalism* (Oxford University Press) p. 130.

Miller, David (1997) On Nationality (Oxford University Press).

Miller, David (2007) *National Responsibility & Global Justice* (Oxford University Press)

Naess, Hans Erik (2019) "Investment Ethics and the global Economy of Sports: The Norwegian Oil Fund, Formula 1 and the 2014 Russian Grand Prix", *Journal of Business Ethics* 158

Nagel, Thomas (2005) "The Problem of Global Justice", *Philosophy & Public Affairs* 33: 113–47.

The Institute for Futures Studies. Working Paper 2021:12

Norges Bank Investment Management (2021), "Government Pension Fund Global", www.nbim.no/en/the-fund/ (accessed May 26, 2021)

Norges Bank Investment Management (2021) "Returns" https://www.nbim.no/en/the-fund/returns/ (accessed May 26, 2021).

Norges Bank Investment Management (2020), "How We Invest", https://www.nbim.no/en/the-fund/how-we-invest/ (accessed May 20, 2021)

Page, Edward (2006) *Climate Change, Justice, and Future Generations* (Edward Elgar Press)

Parfit, Derek (1984) Reasons and Persons (Oxford University Press)

Rawls, John (1999) A Theory of Justice, Rev.Ed. (Oxford University Press)

Vasudeva, Gurneeta (2013) "Weaving Together the Normative and Regulative Roles of Government: How the Norwegian Sovereign Wealth Fund's Responsible Conduct is Shaping Firms' Cross-Border Investments" Organization Science 24(6).

World Population Review (2019) "Median Income by Country", https://worldpopulationreview.com/country-rankings/median-income-bycountry (accessed May 26, 2021)

Paul Bou-Habib¹ and Serena Olsaretti² Children or Migrants as Public Goods?³

Why, and to what extent, must taxpayers share the costs of raising children with parents? The most influential argument over this question has been the *public goods argument*: taxpayers must share costs with parents because and to the extent that child-rearing contributes towards public goods by helping to develop valuable human capital. However, political theorists have not examined the plausibility of the public goods argument in a context in which "replacement immigration" is available: if replacement migration can provide valuable human capital more efficiently than child-rearing, can the public goods argument still justify an obligation for taxpayers to share the costs of child-rearing with parents? This article argues that it can by developing the public goods argument in a new direction that has implications for the fair division of the costs of child-rearing, as well as for other controversies as well.

 $^{^1} Department of Government, University of Essex, pbou@essex.ac.uk$

²ICREA-Universitat Pompeu Fabra, serena.olsaretti@upf.edu

³ This paper will form the basis for the discussion of the chapter on "Parental justice in times of migration and population pressure", in a book manuscript titled *Parental Justice*, under contract with Oxford University Press, and is the product of research undertaken for the project on *Justice and the Family. An Analysis of the Normative Significance of Procreation and Parenthood in a Just Society*, sponsored by an ERC Consolidator Grant (Grant number: 648610; Project Acronym: *Family Justice*).

A central question in the normative debate over the welfare state concerns the fair division of the costs of child-rearing between taxpayers and the family. All sides in this debate agree that taxpayers must support the basic needs of children whenever families are unable to do so. The question is whether families ought to meet a greater share of the costs of child-rearing than they currently do, assuming they are able to do so. The most influential and frequently invoked argument in this debate is the public goods argument, most fully elaborated and defended by Nancy Folbre (1994, 1994b, 2001, 2008) and Rolf George (1987; 1993) and endorsed by a number of other scholars (Alstott, 2005; Anderson, 1999; Esping-Andersen, 2009; Engster, 2007, Rose 2016; Shelby 2016; Strober 2004). It maintains that child-rearing costs must be shared between taxpayers and the family in a way that reflects the fact that childrearing generates human capital from which all citizens benefit, regardless of whether they have children themselves.⁴ The public goods argument plays an important role both in determining the extent to which the state should support individuals in founding and raising a family and as a component of the feminist case for sharing the costs of, given that women do the lion's share of the work in raising children.

Although the public goods argument has been criticised (Rakowski 1991; Casal and Williams 1995; Casal 1999) and defended against criticisms (Olsaretti 2013; Gál, Vanhuysse and Vargha 2018) it has not been closely examined against the assumption that states are able to receive human capital from skilled immigration, or "replacement migration" (UN 2001).⁵ The availability of replacement migration reduces the extent to which citizens need to rely on the family as a source of human capital. This article thus asks: does the possibility of replacement migration weaken or undermine the public goods argument?

This question is likely to become more pressing given two large demographic contexts that will frame and condition the evolution of the welfare state in many societies in the foreseeable future. The first of these contexts is population ageing and the fiscal challenges it poses for the welfare state. The "old-age dependency ratio" of elderly persons to persons of working age is expected to double across most OECD states by 2060, using 2015 as a baseline (OECD 2019, 10). The second is the fact that reversing this ratio by encouraging increases in the fertility-rate is increasingly questioned as a defensible strategy, given the environmental pressures of population growth, especially in advanced economies (Conly 2016; McKibben 1998).

⁴ The language used to convey the public goods argument may inadvertently suggest that children only have instrumental value. However, all proponents of the public goods argument assume that children have the same intrinsic worth that all persons have. Throughout, we have attempted to avoid language that suggests otherwise.

⁵ Nancy Folbre (2008, 178) and Rolf George (1993, 216) do briefly consider the relevance of immigration for the defensibility of family support policies, but not in a sustained fashion.

These evolving contexts make it increasingly important to evaluate the merits of resorting to replacement migration as a source of human capital.

This article defends the public goods argument in favor of family policies in a context in which replacement migration is available. This defence is significant, we believe, not only because it improves our understanding of what is likely to become an increasingly important policy controversy – i.e. the extent of public support for families - but also because it improves our understanding of public good arguments in general, which are frequently used in policy debates. We begin by clarifying the kinds of family policies we focus on in this article. We next explain the concept of "replacement migration" and the reason it poses a challenge for the public goods argument in favor of family polies. The remainder of the article analyses whether this challenge is decisive by pursuing a close discussion of the public goods argument. We distinguish between three different versions of the public goods argument and argue that only one version of that argument- the *pro-natalist* version - is threatened by the availability of replacement migration. Two other versions of the public goods argument – which we call, the *pro-investment* and the *fairness* versions - are not vulnerable to the challenge posed by replacement migration. The article thus shows that the case for family policies remains robust despite the availability of replacement migration in a context of demographic change.

Family Policies

Many public policies indirectly affect how the costs of rearing children are distributed between the taxpayer and parents.⁶ Think of policies regulating the availability of safe and inexpensive public transport which children can use, the existence of a universal and publicly funded health care system, or tax exemptions on food basics, for example. These publicly subsidised amenities and services divert some costs of child-rearing away from the family onto taxpayers. In principle our discussion in what follows bears on the justification of all of these policies. However, we will primarily focus on the policy-packages provided by all welfare states, in some form and to varying degrees, that have an intended and substantial impact on the distribution of the *morally required* costs of raising children – i.e. the costs that must be incurred by *someone* (whether by parents or society at large) in order to give

⁶ By "parents", we refer to adults who have children - typically though not necessarily through procreation - and who raise those children. Many of our claims below apply to adoptive as well as procreative parents. Throughout we talk about "parents" as a group, rather than referring to individual parents. Also, we do not engage with questions that arise because of differences among parents, such as, for example, whether variation in the needs or capacities of their children generates different claims to support for them under the public goods argument. For discussion of these further questions, see Olsaretti (2013) and Wasserman (2017).

children what they have a right to.⁷ These "family policies", as we will call them, include some combination of at least the following three types of provision: (i) publicly-funded parental leave, (ii) free or subsidized pre-school childcare and public schools, and (iii) child tax credits or family allowances. We assume that these policies ought to be generous enough to meet certain recognised benchmarks of children's and parents' needs (e.g. UNICEF 2017).⁸

The family policies we focus on do not distribute certain further costs that are generated when people have and rear children. Our discussion is thus only part of a more comprehensive analysis of the just distribution of the costs of child-rearing. These further costs include some *non-financial* costs of children, of which some may be morally required, e.g., the costs in terms of forgone free time which parents incur in order to care for their children (see Folbre and Bittman 2004; Rose 2016), or the costs in terms of personal autonomy which parents may incur in order to provide their children with continuity of care and stability (Alstott 2005). There are also further costs which arise from people's having children, namely, the costs of children as *added members of society*: these are the costs which children generate as they join the ranks of fellow citizens and lay a claim to their fair shares of social resources. Among these we can include the lifetime environmental cost of an added child, often measured in terms of the child's predicted carbon footprint, which figures prominently in debates over whether the choice to procreate accords with climate justice (see Casal 1999; Young 2001; Heyward 2012; Overall 2012; Conly 2016; McIver 2016; Hickey et al 2016; Cripps 2017).⁹

Our discussion also sets aside the question of what form, exactly, the package of family policies should take. We focus on the justification of the sharing of some of the costs of children between parents and the taxpayer or on what is sometimes referred to as the "collective provision" (Lewis 2003) or the "socialisation" (Olsaretti 2013) of the costs of children – where this is compatible with different views regarding whether family policies that distribute the costs of children should do so in a way that aims at the "defamilialization" of care (e.g. subsidised daycare for babies and toddlers) or at supported familialism (e.g. longer and/or more generous parental leaves) (Saraceno 2010). We also do not discuss further ways in which parents

⁷ These can be contrasted with *morally optional* costs of raising children, e.g. the cost of ensuring that children eat expensive food or travel widely and engage in expensive sports. For a more detailed distinction between "morally required" and "morally optional" costs of child-rearing, see Olsaretti (2018).

⁸ For detailed overviews and analyses of how different welfare state regimes arrange these policies, see Gornick, Meyers and Ross (1999), Lewis (2003), and Saraceno (2010).

⁹ While we do not discuss the environmental costs of children in this article, we assume that any claim we make about the just distribution of the costs of children must be constrained by the demands of a plausible theory of climate justice.

might be benefited – for example, through a pro-parenthood-adjusted pension system (see Sinn 2005; Gál, Vanhuysse and Vargha 2018).

In sum, we will focus on family policies that have an intended and substantial impact on the distribution of the financial costs that are morally required for adequate child-rearing. These family policies currently divide a large proportion of those costs equally between all citizens, regardless of whether they are parents or not, and yet seem primarily to benefit parents and children. This asymmetry is of both practical and theoretical significance. Practically, it makes it important to offer a justification of these policies in the face of what some authors perceive as a "backlash" against them from non-parents (Fineman 2004 and Wolf et al 2011; Burkett 2000). Theoretically, the justification of these policies raises a special challenge, because while many welfare state policies do not benefit all taxpayers equally, this is typically in order to meet needs people are not responsible for having (as in the case of most medical needs or involuntary unemployment). By contrast, it seems harder to justify policies which, while paid for by all citizens, appear to benefit only or mostly some of them as a result of life-plans that they, and not others, embrace (not everyone values parenting), and which people can choose whether or not to pursue (parents generally choose to have children). Family policies thus seem to be under pressure from what political theorists refer to as the challenges of *personal* responsibility and neutrality see (Rakowski 1991; Casal and Williams 1995; Clayton 2006). A successful justification of family policies must explain why taxpayers must share the costs of children with parents, despite the fact that people choose to be parents and even if the state must remain neutral as between the different life-plans that people choose to pursue in life.

The public goods argument is well placed to meet these challenges. Nancy Folbre elaborates the core idea of the argument as follows:

Parents who raise happy, healthy, and successful children create an especially important public good. Children themselves are not the only beneficiaries. Employers profit from access to productive workers. The elderly benefit from Social Security taxes paid by the younger generation...Fellow citizens gain from having productive and law-abiding neighbors. (2001, 50).¹⁰

By emphasising the contribution of parenting rather than the interest in parenting, the public goods argument promises to meet head-on the challenges of personal

¹⁰ While much can be said about whether these goods are appropriately characterized as *public* goods in the technical sense used by economists (i.e. they are non-excludable and non-rival), as opposed to what some have called "normative public goods" (White 2003) or "socialised goods" (Olsaretti 2013), and whether and why this matters, we do not need to take up these questions here.

responsibility and neutrality mentioned earlier. In response to the personal responsibility challenge, the public goods argument can say that people must be required to bear the costs of their choices only if their choices would otherwise generate costs for others, as is the case, for example, when an imprudent motorcyclist prefers to ride without a helmet, and potentially creates costs for others in having to provide him with emergency healthcare (Fleurbaey 1995). Because parents benefit society at large by having and rearing children, the challenge from personal responsibility is mistargeted when it is directed against parents (see Casal and Williams 1995).

In response to the neutrality challenge, the public goods argument can offer the following rejoinder. Because the basis for sharing child-rearing costs with parents is the contribution they make through child-rearing, rather than their interests in parenting, cost-sharing does not presume any controversial judgement that parenting is an intrinsic part of a good life. True enough, the concern with state neutrality has implications regarding which of the various benefits that parents create are relevant for justifying cost-sharing with them. Cost-sharing with parents could not be justified, compatibly with neutrality, if child-rearing generated benefits that are regarded as such only from the perspective of one or another controversial concepttion of the good life. It is important to observe, therefore, that child- rearing creates benefits which all, if not most, members of society can reasonably be presumed to want (these are what John Rawls famously calls "social primary goods" (1999, 54)). One of the central benefits of demographic renewal – the maintenance of or enlargement of the tax base – is an uncontroversial benefit from the perspective of most conceptions of the good life.

The Challenge of Replacement Migration

So-called "replacement migration" can contribute some of the very same public goods that parents contribute by having and rearing children. The question we address in the remainder of the article is this: *assuming* that replacement migration does provide the same public goods that figure in the public goods argument for sharing the costs of children with parents, does the availability of replacement migration undermine the public goods argument? This challenge has received barely any attention in the literature (although see Folbre 2008, 279; and George 1993, 216 for brief references to it). In this section we begin by clarifying the challenge.¹¹

¹¹ For a thought-provoking discussion of the related but more general question, of *how much* of the cost of child-rearing others are required to share with parents, see Shields (2021).

The term "replacement migration" originated in a 2001 report issued by the United Nations Population Division.¹² It refers to immigration that makes up for a demographic deficit, where a "demographic deficit" is a deficit either in a country's population as a whole, or in a particular group of persons within its population (e.g. working-age adults). Different kinds of "replacement migration" are thus possible, depending on the demographic deficit it replaces.¹³ In this article, we focus on replacement migration that aims to replace a deficit in skilled workers because this kind of replacement migration substitutes most closely for the public goods that parents are said to contribute in raising children, and thus constitutes the most serious challenge to the public goods argument for cost-sharing with parents. Different policies can be used to facilitate this kind of replacement migration. Countries can adopt a points-based immigration system that issues visas to immigrants according to their skills (e.g. Australia and Canada), or an employer-based immigration system that issues visas to immigrants who have accepted job-offers within a certain range of occupations (e.g. USA). These policies can be adjusted in order to attract immigrants with certain kinds of skills and who belong to certain age-groups, and they can vary in their specific provisions, e.g. the extent to which they permit immigrants to bring relatives with them (Gosseries and Zwarthoed 2017; Bou-Habib 2018.)

A simple example helps to illustrate the challenge that replacement migration poses for the public goods argument. In 1996, some 254,000 children were born in Australia (ABS 1996, 10). In 2014, when this cohort turned 18 years of age, Australia admitted around 125,000 skilled immigrants (DIBP 2014, 4) The Australian government could have influenced the sizes of these two groups of persons through various policy avenues. It could have offered less generous public support for families just prior to 1996, which would have discouraged some families from having as many children, and it could have adjusted its immigration policy, so as to increase the number of skilled immigrants entering in 2014. Stated more generally, the government could have shifted, in some degree, from relying on procreation to relying on immigration as a source of human capital. (This is, of course, a recurring possibility: the Australian government, like almost every government, can affect the balance between these two sources of human capital today and in the future.) Now assume that the cost for Australian taxpayers of obtaining human capital *via* pro-

 $^{^{12}}$ See UN (2001). Our usage of "replacement migration" in this article is in line with how the term is used that report.

¹³ The 2001 UN report estimated the number of immigrants that various low-fertility countries needed in order to make up for various kinds of demographic deficit, including population size, and the ratio of working adults to retirees.

creation. This assumption is plausible: replacement migration is certainly not costless, to be sure, but it is doubtful that the average public cost of integrating a skilled migrant into the labor force exceeds the average public cost involved in the rearing of children from birth to adulthood (cf. Bou-Habib 2018). A critic of family policies could now say that the public goods argument cannot justify an obligation for Australian taxpayers to share the costs of child-rearing with Australian parents, at least for any amount of human capital that Australia could obtain *via* replacement migration. For there is, presumably, no obligation for taxpayers to share costs with producers of a given public good if the public good in question can be obtained at lesser cost in some other way (cf. Shields 2021). Thus, to the extent that replacement migration is possible, the public goods argument for cost-sharing with parents seems to fail.

The replacement migration challenge applies regardless of whether or not there are concerns about overpopulation, but it is worth nothing that the latter concerns seem to reinforce it. As one writer notes, "[i]n an age of actual or prospective overpopulation..... the suggestion that an allowance for children is justified by parents' serving a societal function" is "entirely unpersuasive" (Brazer 1977, cited in George 1993, 216). Indeed, in a context of global overpopulation, parents may be argued to produce, instead of a public good, a public *bad* (Casal 1999).¹⁴ We believe our defence of the public goods argument against the replacement migration challenge holds, even once that challenge is reinforced by concerns about overpopulation and we briefly indicate why this is so, after fully stating our defence (see fn. 22 below). However, given the complexities involved, a detailed discussion of how concerns about overpopulation bear on our defence must await a separate discussion.

Three Versions of the Public Goods Argument

To assess the replacement migration challenge, we now suggest a three-fold distinction between different versions of the public goods argument. These three versions of the argument are generally collapsed in the literature.¹⁵ Disaggregating the public goods argument is important because this enables us to determine whether it is vulnerable to the replacement migration challenge in any or all of its possible versions.

On the first, *pro-natalist* version of the public goods argument, family policies are justified insofar as they are necessary or effective means for incentivising people *to*

¹⁴ Folbre (1994, 2008) also anticipates the concern with overpopulation as a possible objection to the public goods argument.

¹⁵ An exception is Casal and Williams (1995), who distinguish between our first two versions of the public goods argument, on the one hand, and the third version, on the other.

have children. If families produce the important public goods mentioned earlier by having children, and if it is true that, absent public support for the costs of childrearing, people have too few children, then, given independent, moral or prudential reasons that we have, as a society, to ensure that public goods of this sort are produced, family policies that enable this to happen are justified. We call this version of the public goods argument "pro-natalist" because it emphasises boosting the fertility-rate as an aim that justifies family policies.

A second, *pro-investment* version of the public goods argument points to the importance of supporting a certain kind of parental investment in children, rather than supporting procreation. Here, unlike the pro-natalist argument, it is assumed that adults do have enough children (fertility rates are not what is in question), and that there are public goods-based reasons to ensure that children are raised as productive members of society. The pro-investment version of the public goods argument assumes that investment in the family is necessary or effective for ensuring that children are raised in this way. It allows that in the absence of public support for families, people will continue to have children. Where it differs from the pro-natalist version of the public goods argument is that its aim is to ensure *effective parenting* rather than a higher fertility rate (see, for example, Heckman and Masterov 2007; Heckman 2013; Putman 2015).

Note that both the pro-natalist and the pro-investment versions of the public goods argument are *forward-looking*: they justify sharing the costs of children by pointing to certain further consequences that sharing the costs of children is predicted to bring about. By contrast, the third version of the public goods argument is *backward-looking*, in the sense that it justifies sharing the costs of children by reference to what we owe parents. We call this the *fairness version* of the public goods argument. It states that family policies that share costs of child-rearing are justified as a way of achieving a fair distribution of the burdens as well as of the benefits of child-rearing between parents and others. The argument rests on two main premises. The first, empirical premise, which we have already stated a number of times, is that by having and raising children, parents generate valuable human capital for a society from which all citizens benefit. The second, normative premise is a principle of distributive fairness.¹⁶ This demands that some of the burdens, as

¹⁶ A full elaboration of the fairness version of the public goods argument would specify what exactly constitutes fair burden sharing among producers and beneficiaries of a public good. The two main views of this are (a) the *equal ratio* view – i.e., that beneficiaries must contribute towards the costs of production in proportion to the share of benefits they obtain (Nagel and Murphy 2004) and (b) the *equal net benefit* view – i.e., that beneficiaries must share costs with producers so that they and the producers end up as equal net beneficiaries of public goods-producing activities (Miller and Taylor 2018). Nothing in our discussion below depends on which of these views (or other views of fair burden sharing) is justified.

well as the benefits, of socially necessary activities be shared between those who undertake them and their beneficiaries, so that no unfair freeriding takes place.¹⁷ We do not defend here the normative premise of the fairness version of the public goods argument; this brief description of it should serve our purposes for now.¹⁸

Having distinguished three versions of the public goods argument, we now briefly assess the replacement migration challenge for the first two, forwardlooking versions of the argument. The implications of the replacement migration for the fairness version of the argument require a more detailed discussion, which we turn to in the following section.

It is relatively clear that the *pro-natalist* version of the public goods argument is undermined by the possibility of replacement migration. As Bou-Habib (2018) argues, if the inclusion in our societies of migrant adults, including migrant adults who come with children, can offset the need to increase the local fertility rate, then the pro-natalist version will not support sharing the costs of child-rearing.

By contrast, the *pro-investment* version of the public goods argument *can* survive even in the face of replacement migration. This is true, recall, so long as three conditions hold: (i) our fellow citizens will continue to have children, with or without the support of family policies (fertility rates are not what is in question), (ii) there are reasons (both of childhood justice and in terms of what is in the common interest) to ensure that children are raised in certain ways (e.g. not in poverty; as productive members of society; as law-abiding citizens), and (iii) public support for parents is effective for ensuring that most or all children are raised in these ways.

However, while the pro-investment version of the incentive argument is not undermined by replacement migration, this version of the argument is limited in two respects. First, it only grounds public support for low-income parents, given that wealthier parents can already afford to cover the morally required costs of children – and typically do already incur substantial expenses to raise children (Putman 2015; Burggraf 1993). Thus while the pro- investment version of the public goods argument survives the challenge posed by replacement migration, this is small comfort for many proponents of the public goods argument, given that it does not yield a case for publicly funded parental leave for all parents, nor for publicly funded universal health care, childcare and education for all children, but only for

¹⁷ By "socially necessary" activities we refer those activities that help maintain society as an ongoing and just scheme of social cooperation. For example, a society would not continue without demographic renewal, and it could not be just if individuals did not comply with a system of shared rules that work to the benefit of all. Child-rearing and complying with a system of shared rules are then both "socially necessary" activities as we use that term.

¹⁸ One frequently adduced defense of the normative principle is the so-called "principle of fair play" first set out by Hart (1955) and Rawls (1968; 1971) and elaborated in many directions by subsequent scholars (Klosko 1987; Cullity 1995; Simmons 2001). We do not have space here to discuss this principle.

children in low-income households. Second, and furthermore, low-income parents would only be entitled to the benefits of family policies *assuming* that other means of escaping poverty (e.g. through effective welfare-to-work policies) are not available or are less effective than family policies in ensuring that their children are adequately raised. In summary, the pro-investment version of the public goods argument has application to a narrow context; it will not necessarily support the universal provision of the kinds of family policies we discussed in the previous section.

Whether or not we believe that this fact constitutes a point against the proinvestment version (whether or not, that is, we have antecedent convictions that family policies should be available for *all* parents rather than just low-income parents), we believe the first two forward-looking versions of the public goods argument do not, in any case, fully capture the underlying convictions that move people to embrace the public goods argument. To see why, imagine that those who have children go to great lengths, even in the absence of family policies, to ensure that their children are duly cared for, have adequate healthcare and receive a good education; and that in so doing, in addition to benefiting their children, they thereby help to produce essential public goods for others in society. (In other words, imagine that family policies are not *necessary* to induce parents to have children and to invest in raising them well.) A society in which parents are made to pay entirely for the costs of having and raising children, while at the same time ensuring that those public goods are shared among everyone, seems, on its face, to distribute the burdens and benefits of social cooperation unfairly, and to the detriment of the "producers", in this case parents (cf. Olsaretti 2013). It is thus particularly important to consider whether the replacement migration challenge undermines the fairness version of the public goods argument.

Fairness and Public Goods: A Question of Baselines

The availability of replacement migration poses a challenge to the fairness version of the public goods argument because that argument assumes that those who are said to incur obligations to do so as a result of *benefiting* from the producers' activities. To the extent that some replacement migration is available as a more cost-effective source of human capital, it could be said that the fact that parents have and rear children at their current rate (say, 1.55 children per woman, to use the 2018 average European fertility rate)¹⁹ is no longer beneficial to society at large, or no longer beneficial enough to justify sharing the costs of all children.

¹⁹ See the Eurostat "Fertility Statistics" available at https://ec.europa.eu/eurostat/statistics-explained/index.php/Fertility_statistics.

The following parallel helps to illustrate this challenge. Suppose a society is deciding on a policy to reach its collective emissions target. It can meet the target in two ways: (a) by subsidizing weekday train-travel or (b) by subsidizing only weekend train-travel but then importing some amount of cleanly produced energy from a neighboring country. Furthermore, suppose that policy (b) is cheaper than policy (a). Under these circumstances, weekday train commuting does not benefit society relative to the available alternative, and it might then be thought unreasonable for weekday train-commuters to insist that taxpayers subsidize some of the costs of their weekday train-travel. Similarly, if the benefits that parents provide by rearing children can be provided by replacement migration in a more cost- effective way, this seems to challenge the fairness-based case for public support for parents. Parents do not benefit others in society relative to the available alternative of replacement migration. In this scenario, it would thus appear that appealing to the fact that parents are producing benefits for society at large as a reason to share costs with them is out of place. Rolf George (who opposes this line of thinking) puts the point in a deliberately blunt way: "Why should one pay for locally produced children if they can be obtained so much more cheaply by import?" (1993, 216; cf. Shields 2021). More precisely, if it were successful, this challenge to the fairness version of the public goods argument would establish the following: Parents do not have claims to having the costs of chid-rearing shared by others beyond the level at which others would begin benefitting less from parenting than they would from available replacement migration.

To defend the public goods argument against the challenge of replacement migration, we now bring to view and address a crucial question that, to our knowledge, discussions of public goods arguments have neglected. This is the question of the baseline by reference to which we should judge whether other people's activities benefit us such that we incur obligations to share costs with them (as the public goods argument maintains).²⁰ We show that the challenge from replacement migration relies on an implausible view of this baseline, and offer a more defensible view of it, in light of which cost-sharing with parents is justified despite the availability of replacement migration.

To appreciate the kind of baseline that is assumed by the replacement migration challenge, note that replacement migration cannot under any realistic scenario fully

²⁰ The question of the relevant "baseline" for comparison arises in many discussions in political theory in which a decision needs to be made about the moral consequences that follow when some persons affect the outcomes of others. These include discussions about the nature of "harm" (Bradley 2012), "coercion" (Nozick, 1969), and "exploitation" (Wertheimer, 1999). We are suggesting that the question of the baseline is also relevant in public goods arguments; this is not surprising, because whether someone "benefits" others is yet another instance of action that affects others that can have important moral consequences.

supplant a society's reliance on local child-rearing; the continuous renewal of a population is only possible if at least some local people continue to have and rear children. The replacement migration challenge is thus not that local child-rearing is completely dispensable, but that the number of children local parents have exceeds some desirable level. More specifically, the challenge assumes what we will call an optimal baseline view: this is the view that producers only "benefit" the beneficiaries if they produce up to and no more than the amount that is optimal for the beneficiaries. The number of children reared beyond that optimal level are no longer public goods: when parents engage in this additional child-rearing they are, in effect, "oversupplying" their services, such that, on balance (i.e. taking into account the costs of children as well as their benefits), the additional children constitute negative externalities, rather than public goods. To illustrate: if the availability of replacement migration meant that the desirable fertility rate, as judged by the lights of the optimal baseline, were one child per woman, then the public goods argument – when it employs the optimal baseline view - would not justify sharing the costs of any children in excess of that amount, and family policies should be adjusted to reflect that fact, for example, by only offering child tax credit or publicly funded parental leave for one child per woman.

Having clarified that the replacement migration assumes an optimal baseline view, our response to the challenge now proceeds in three steps.

The first step observes that the optimal baseline is not the only possible baseline that a public goods argument can employ. Consider an alternative view of the baseline, which we will call the *no production baseline view*. It says that we should judge whether others are benefitted by child-rearing by asking whether they are made better off by it than they would be if *no new* children were born. If the public goods argument employed the no production baseline, others would have obligations of fairness to parents to share costs of child-rearing with them for *whatever* number of children they have, provided only that that number of children makes them better off compared to no new children. Depending on whether it employs the optimal baseline or the no production baseline, the public goods argument thus yields very different implications for when others must share the costs of childrearing with parents.

Our second step argues that the optimal baseline view, which underpins the replacement migration challenge, is implausible for the context of child-rearing. The optimal baseline view is, to be sure, a highly plausible view for other contexts. Consider, for example the provision of a paradigmatic public good, namely national defence. If weapons factories were to produce military equipment in excess of the optimal amount needed by a society, it seems right that taxpayers do not have an obligation to shoulder the costs of producing the oversupply. By contrast, our con-

tention is that the optimal baseline is implausible when the activity that generates public goods is child-rearing. To see why these different judgements seem justified, it is helpful to imagine a proposal that will strike many as unacceptable (and rightly so, in our view). Suppose that everyone in a given society would be better off if the fertility rate were higher than it is. In view of this, the proposal is made that people who do not have children should be held liable for the under-supply of children on the grounds that their procreative choices (i.e. in this case, their choices to not have children) make others worse off than they would be relative to the optimal number of children.²¹ We can suppose that people who do not have children in this society would be held liable for this by being deprived of some of the benefits others bring about by rearing children, such as access to publicly funded pensions. We believe most people would find this proposal unacceptable (for a contrary view, see Sinn 2005).

The reason for this, we submit, is that everyone has a compelling interest in being able to decide whether or not to have children, and should not be penalised for pursuing this compelling interest in one way or another, provided they do not violate anyone's rights. Note that we are not claiming that people may not be denied economic benefits they need to pursue their compelling interests. Although people have, we assume, a compelling interest in freedom of conscience, others are not required to share with them the costs of building their churches, for example. But provided that they exercise their freedom of conscience in ways that respect other people's rights, they may not be charged, say, higher taxes for exercising it in one rather than another way, only because fellow citizens would benefit from an alternative exercise of their freedom of conscience. Similarly, people may not be deprived of publicly funded pensions just because their exercise of procreative choice fails to yield the optimal number of children for a society. Thus, whether the compelling interests of people are implicated in the activities that produce benefits matters in identifying which baseline we should adopt when determining costsharing with them: their compelling interests set the parameters for how much we can legitimately expect to be benefited from their activities before we must share with them the costs of their productive activities.

We define "compelling interests" in line with John Rawls' account, as interests that are of central importance to all citizens, despite their having different conceptions of the good life, and that merit the protection afforded by a principle of equal basic liberty. On Rawls' well-known view, those interests – including the interest in freedom of speech, conscience, and association, and occupational choice – are ones

²¹ See Mulgan (2006) for a critique of utilitarian moral theories for entailing that there is a moral duty to have children. For the view that that there may be such a duty under some circumstances, see Smilansky (1995) and Gheaus (2015).

which people have in virtue of being "moral persons", that is, persons with the capacity to develop and pursue their own plans of life and a sense of justice (1996, 293). For Rawls, the protection of these interests has priority over socio-economic justice, at least once a society reaches a certain minimal threshold of economic wellbeing (1999, 266). The priority of freedom of occupational choice over claims of justice to income and wealth entails, for example, that we may not choose a regime in which people are coerced to work at their most productive occupation in order to maximally improve the prospects of the least well off. We claim that the interest in procreative choice – understood as the freedom to choose whether or not to bring into the world and parent at least one child counts as a compelling interests by the standard Rawls adopts.²² That being said, we do not assume that only the interests. It is possible that the latter include further, very weighty interests of persons.

Three brief points need to be emphasised in regard to the compelling interest in procreative choice. First, the interest that we claim is of central importance to all citizens is not an interest in *procreation*, but in procreative *choice*. Second, we do not assume (indeed, we deny) that procreative choice is of greater value or importance to citizens than the choice to parent by adoption. Third, we will not attempt to state the number of children persons must be able to have in order for their interest in procreative choice to be satisfied. This is a complex and profound question in ethics (McKibben 1998; Overall 2012; Conly 2015). We assume that there is an upper threshold beyond which parents cannot reasonably claim that having additional children is still of central importance to their lives, or at least as great in importance as is the choice to have just one or two children. However, for lack of space, we cannot explore the arguments that would need to be made for specifying such a threshold and we acknowledge that the implications of our argument in this article are sensitive to this issue.

To see the relevance of construing procreative choice as a compelling interest, consider the following analogy to the compelling interests people have in freedom of occupational choice. Suppose it were suggested that we regard the tax-revenue that the state could obtain if everyone worked at their most productive occupation as the baseline relative to which we should judge whether individuals are benefiting others in their occupational choices. On this view, someone who chose to work as a teacher, say, but who could have worked more productively as a lawyer, would count as producing a negative externality for which he may be held liable (this might, for

²² Rawls does not list procreative liberty as one of the basic liberties; while noting his injunction that the list of basic liberties should be limited (1996, 296), we believe this is an oversight on Rawls' part. Note that the right to family life, including the right to found a family, is proclaimed in Article 16 of the Universal Declaration of Human Rights.

example, be reflected in his tax liability). This, we submit, is unacceptable for the reason that Rawls gives. Freedom of occupational choice, as a basic liberty, protects the teacher against liability for pursuing a less than optimally productive occupation. The same should be said of procreative choice: it, too, protects parents against liability for having fewer or more than the optimally productive number of children. The optimal baseline is therefore not plausible for determining whether and when taxpayers must share the costs of child-rearing with parents. It *is* plausible, by contrast, in our earlier national defence example, because no compelling interest is exercised in the course of producing military equipment. In that context, it *is* reasonable to hold people liable for over- supplying (or under-supplying) relative to the optimal amount that a society needs.

In a third and final step, we now explain the most plausible baseline view for the public goods argument when it is used for the context of child-rearing. Note, to begin with, that we should reject what we earlier called the "no production baseline" view. According to the no production baseline view, beneficiaries must share costs with producers so long as beneficiaries benefit *to any extent*, no matter how minimal, relative to the absence of production altogether (e.g. compared to how society would fare if no new children were brought into existence). The example of weapons production baseline view. If weapons' producers produce in excess of the optimal amount of weapons, they cannot reasonably expect taxpayers to share with them the costs of that oversupply just because an excess of weapons is at least better than no weapons.

We can state the most plausible baseline view if we first observe the common problem that the optimal and the no production baseline views share: neither view is appropriately sensitive to the fact that compelling interests are sometimes implicated in the activity that produces public goods. To employ the optimal baseline view across-the-board – that is, for all activities that produce public goods – is to overlook the fact that compelling interests are implicated in some of those activities and that those compelling interests set limits to the maximising demands made by the optimal baseline view (this is the case, we believe, when it comes to child-rearing). The no production baseline view, for its part, overlooks the fact that producers of public goods can only reasonably expect others to share with them the costs of non-optimal production if their productive activities are necessary for satisfying a compelling interest (that is why weapons producers cannot reasonably expect cost-sharing for an oversupply of weapons).

We therefore propose that public goods arguments must employ what we call the *interest-sensitive baseline view*, so named because it selects the relevant baseline for establishing cost-sharing obligations with producers according to whether a com-

pelling interest is, or is not, implicated in the activity that produces the public good in question. If, on the one hand, no compelling interest is implicated in the activity that produces a public good, the correct baseline that triggers an obligation to share costs with producers is the optimal baseline (so, this is the correct baseline for producing weapons for national defence). If, on the other hand, a compelling interest is implicated in the activity that produces a public good, the correct baseline is more complex. Beneficiaries have an obligation to share costs with producers if they benefit relative to the absence of the productive activity and are brought as close as possible to the optimal baseline as is compatible with the producers' compelling interest being respected. If beneficiaries either do not benefit relative to the absence of that activity, or if the productive activity stretches beyond what is necessary for securing the relevant compelling interest (so that the departure from the optimal baseline is not required to satisfy the compelling interest), no obligation to share costs with producers arises.²³ Because the interest in procreative choice is a compelling interest, this complex baseline is the correct baseline to use for determining whether taxpayers must share the costs of child-rearing with parents.

If, as we have argued, the public goods argument should employ the interestsensitive baseline, then the threat posed by replacement migration to the justification of family policies is significantly diminished. That challenge assumes, recall, that family policies that support parents are justified only for that number of children which is optimal - which, given the availability of replacement migration, may be much lower than the number of children parents would like to have. This assumption, we have argued, is mistaken. Family policies are justified even for children "in excess" of the optimal baseline, if it is true that parenting those children still benefits others in society relative to the absence of those children altogether and that number of children is necessary for everyone's exercising their compelling interest in procreative choice. So, while we have not sought to specify an upper threshold of children that parents must be able to have in order to satisfy that compelling interest, we can at least establish the following general conclusion: family policies that support parents who rear children below this upper threshold are justified, so long as parents thereby benefit others relative to the absence of those children. We believe this is likely to be true under realistic conditions in many scenarios, even when we take into account concerns about overpopulation.²⁴

²³ In this scenario, some other argument might still be able to support an obligation for taxpayers to share the costs of child-rearing with parents. What distinguishes the public goods argument from a "pure" interest-based case for sharing the costs of children should be apparent here. Unless condition (i) is satisfied, no public goods argument exists; the appeal to compelling interests, in our view, only enters into the argument by constraining how much we can legitimately expect to be benefited by others' productive activities.

²⁴ As we mentioned earlier, the replacement migration challenge to family policies is reinforced by

Conclusion

The public goods argument is the most influential argument in the long-running debate amongst political theorists and social scientists over how the costs of child-rearing should be shared between the taxpayer and the family. Yet it has not been examined under the assumption that replacement migration is available as an alternative source of the public goods that child-rearing provides. To assess its robustness under this assumption, we have developed the public goods argument in a new direction which emphasises the relevance of the fact that the activities through which we produce public goods sometimes involve our most compelling interests. As well as improving our understanding of the normative underpinnings of family policies, we believe our discussion can improve our understanding of other controversies in which the public goods argument plays a central role. We wish to close with one example that illustrates 1the broader relevance of our discussion.

Consider the frequently made claim that higher education produces public goods for everyone in society – a well-educated citizenry is of vital importance for the health of democratic institutions, as well as for a strong economy - and that the state should therefore share the costs of higher education with students (Barr and Crawford 2005; Martin 2017). Suppose someone were to argue that the state need share the costs only of that kind and level of higher education that is optimal for producing public goods. It might be the case, for instance, that students would contribute the optimal amount of public goods if they shifted, in some number, from pursuing degrees in Humanities subjects, towards degrees in STEM subjects (science, technology, engineering and mathematics). Our discussion of the public goods argument suggests that we must ask whether citizens have a compelling interest in being able to choose the kind of higher education they should pursue. If they do, it might be unfair for the state to support only those degrees in the Humanities that are needed for producing an optimal amount of public goods from higher education and to ask students to privately fund any Humanities degrees beyond that number. Here, too, as in the case of child- rearing, the fact that a benefits-producing activity involves the exercise of a compelling interest matters; the public goods argument can, and must, do justice to that interest, as well as to the interests of beneficiaries in receiving the public goods.

concerns about overpopulation, but if the relevant baseline for assessing whether parents make a social contribution is the interest-sensitive baseline that we have defended, then child-rearing may well make such a contribution, even once we take on board concerns about overpopulation.

References

ABS (Australia Bureau of Statistics). 1996. *Births: Australia*. Canberra: ABS Catalogue No 3301.0.

Alstott, Anne. 2005. *No Exit: What Parents Owe their Children and What Society Owes Parents*. Oxford: Oxford University Press.

Anderson, Elizabeth. 1999. "What's the Point of Equality?" Ethics 109: 287-337.

Barr, Nicholas. and Ian Crawford. 2005. *Financing Higher Education*. London: Routledge. Bou-Habib, Paul. 2019. "The Case for Replacement Migration." *Journal of Political Philosophy* 27: 67–87.

Bradley, Ben. 2012. "Doing Away with Harm." *Philosophy and Phenomenological Research* 85: 390–412.

Brazer, Harvey. 1977. "Comments.". In J. Pechman, ed.), *Comprehensive Income Taxation*. Washington, D.C.: Brookings Institute.

Burggraf, Shirley. 1993. "How Should the Costs of Child Rearing Be Distributed?" *Challenge* 36: 48–55.

Burkett, Elinor. 2000. The Baby Boon. New York, NY: The Free Press.

Casal, Paula. 1999. "Environmentalism, Procreation, and the Principe of Fairness." *Public Affairs Quarterly* 13: 363–376.

Casal, Paula. and Andrew. Williams. 1995. "Rights, Equality and Procreation." *Analyse & Kritik* 17: 93–116.

Clayton, Matthew. 2006. *Justice and Legitimacy in Upbringing*. Oxford: Oxford University Press.

Conly, Sarah. 2016. *One Child. Do We Have a Right to More?* Oxford: Oxford University Press.

Cripps, Elizabeth. 2017. "Do Parents Have a Special Duty to Mitigate Climate Change?" *Politics, Philosophy & Economics* 16: 308–325.

Cullity, Garrett. 1995. "Moral Free Riding". *Philosophy and Public Affairs* 24: 3–34. DIBP (Australia). 2014. *2014-15 Migration Programme Report*. Available at: https://www.homeaffairs.gov.au/research-and-statistics/statistics/visastatistics/live/migration-program (accessed February 12, 2021). Engster, Daniel. 2008. *The Heart of Justice*. Oxford: Oxford University Press. Esping-Andersen, Gøsta. 2009. *The Incomplete Revolution*. Cambridge: Polity Press.

Fleurbaey, Marc. 1995. "Equal Opportunity or Equal Social Outcome." *Economics and Philosophy* 11: 25–55.

Folbre, Nancy. 1994. "Children as Public Goods." *American Economic Review* 84: 86–90.

Folbre, Nancy. 1994b. Who Pays for the Kids? London: Routledge.

Folbre, Nancy. 2001. *The Invisible Heart. Economics and Family Values*. New York: The New Press.

Folbre, Nancy. 2008. *Valuing Children: Rethinking the Economics of the Family*. Cambridge, MA: Harvard University Press.

Folbre, Nancy and Michael Bittman, eds. 2004. *Family Time: The Social Organization of Care*. London: Routledge.

Gál, Vanhuysse and Vargha 2018. "Pro-Elderly Welfare States within Child-Oriented Societies." *Journal of European Public Policy* 25: 944–958.

George, Rolf. 1987. "Who Should Bear the Costs of Children?" *Public Affairs Quarterly* 1: 1–42.

George, Rolf. 1993. "On the External Benefits of Children." In Diana Meyers, Kenneth Kipnis, and Cornelius Murphy, eds., *Kindred Matters. Rethinking the Philosophy of the Family*. Ithaca, NY: Cornell University Press, 209–217.

Gheaus, Anca. 2016. "Could There Ever Be a Duty to Have Children?" In Sarah Hannan, Samantha Brennan, and Richard Vernon, eds., *Permissible Progeny?* Oxford: Oxford University Press, 87–106.

Gornick, Janet., Marcia Meyers, and Katherin Ross. 1999. "Public Childcare, Parental Leave, and Employment." In Daine Sainsbury, ed., *Gender and Welfare State Regimes*. Oxford: Oxford University Press, 117–146.

Gosseries, Axel and Danielle Zwarthoed. 2017. "Generations and Global Justice." In David Held and Pietro Maffetone, eds. *Global Political Theory*. Cambridge: Polity Press, 281–304.

Hart, H. L. A. 1955. "Are There Any Natural Rights?" *The Philosophical Review* 64: 175–191.

Heckman, James. 2013. Giving Kids a Fair Chance. Boston: MIT Press.

Heckman, James. and Dimitriy Masterov. 2007. "The Productivity Argument for Investing in Young Children." *Review of Agricultural Economics* 29: 446–493.

Heyward, Clare 2012. "A Growing Problem? Dealing with Population Increases in Climate Justice." *Ethical Perspectives* 19: 703–732.

Hickey, Colin, Travis Rieder, and Jake Earl. 2016. "Population Engineering and the Fight Against Climate Change." *Social Theory and Practice* 42: 845–870.

Klosko, George. 1987. "Presumptive Benefit, Fairness and Political Obligation, *Philosophy and Public Affairs* 16: 241–259.

Lewis, Jane 2003. *Should We Worry About Family Change*? Toronto: University of Toronto Press.

Martin. Christopher 2017. "Should the Public Pay for Higher Education? Education, Liberty, and Educational Debt." *Theory and Research in Education* 15: 38–52.

McIver, Corey 2016. "Procreation or Appropriation?" In Samantha Brennan and S. Hannan, eds., In Sarah Hannan, Samantha Brennan, and Richard Vernon, eds., *Permissible Progeny?* Oxford: Oxford University Press, 107–128.

McKibben, Bill 1998. Maybe One. New York, NY: Simon & Schuster.

Miller, David and Isaac Taylor. 2018. "Public Goods." In S. Olsaretti, ed., *The Oxford Handbook of Distributive Justice*. Oxford: Oxford University Press, 556–575.

Mulgan, Tim. 2001. *The Demands of Consequentialism*. Oxford: Oxford University Press. Nagel, Thomas and Liam Murphy. 2004. *The Myth of Ownership: Taxes and Justice*. Oxford: Oxford University Press.

Nozick, Robert 1969. "Coercion." In S. Morgenbesser, P. Suppes, and M. White, eds., *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*, New York: St. Martin's Press, 440–472.

OECD. 2019. *Fiscal Challenges and Inclusive Growth in Ageing Societies*. Paris: OECD Economic Policy Papers.

Olsaretti, Serena 2013. "Children as Public Goods?" *Philosophy and Public Affairs* 41: 226–258.

Olsaretti, Serena 2018. "The costs of children." In A. Gheaus, G. Calder and J. De Wispelaere, eds., *The Routledge Handbook of the Philosophy of Childhood and Children*. Abingdon-on-Thames, UK: Routledge, 339–350.

Overall, Christine 2012. *Why Have Children? The Ethical Debate*. Cambridge, MA: MIT Press.

Putnam, Robert 2016. *Our Kids: The American Dream in Crisis*. New York, NY: Simon & Schuster.

Rakowski, Eric. 2003. Equal Justice. Oxford: Clarendon Press.

Rawls, John 1996. *Political Liberalism*. New York, NY: Columbia University Press. Rawls, John 1999. *A Theory of Justice (Revised Edition)*. Cambridge, MA: Harvard University Press.

Rawls, John 2001. *Justice as Fairness: A Restatement*. Cambridge, MA: Harvard University Press.

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

Saraceno, Chiara 2010. "Can We Identify Intergenerational Policy Regimes in Europe?" *European Societies* 12: 675–696.

Shelby, Tommie 2016. *Dark Ghettoes: Injustice, Desert and Reform*. Cambridge, MA: Harvard University Press.

Shields, Liam 2021. "Children as Public Goods: At What Cost?" Unpublished manuscript.

Simmons, A. John. "Fair Play and Political Obligation: Twenty Years Later". In A. John Simmons, *Justification and Legitimacy*, Cambridge: Cambridge University Press, 27–42.

Sinn, Hans-Werner 2005. "Europe's Demographic Deficit: A Plea for a Child Pension System." *De Economist* 153: 1–45.

Smilanksy, Saul 1995. "Is There a Moral Obligation to Have Children?" *Journal of Applied Philosophy* 12: 41–53.

Strober, Myra 2004. "Children as a Public Good." Dissent 51: 57-61.

UN. 2001, *Replacement Migration: Is it a Solution to Declining and Ageing Populations*. New York: United Nations Publication.

UNICEF. 2017. *Building the Future: Children and the Sustainable Development Goals in Rich Countries.* Florence, Italy: UNICEF Office of Research.

Wasserman. David. 2017. "Justice, Procreation, and the Costs of Having and Raising Disabled Children." In L. Francis, *The Oxford Handbook of Reproductive Ethics*, Oxford: Oxford University Press, 464–477.

Wertheimer, Alan. 1999. *Exploitation*. Princeton: Princeton University Press. White, Stuart 2003. *The Civic Minimum*. Oxford: Oxford University Press.

Wolf, Douglas et al 2011. "Fiscal Externalities of Becoming a Parent." *Population Development Review* 27: 241–266.

Young, Thomas 2001. "Overconsumption and Procreation: Are they Morally Equivalent?" *Journal of Applied Philosophy* 18 (2): 183–192.
Tim Campbell,¹ Martin Kolk,^{1,2} Julia Mosquera¹ Universal Procreation Rights and Future Generations³

Do we have demands of procreative justice towards future generations? And if so, how can we ensure that such demands are met? It is often acknowledged that the policies implemented by governments can constrain people's procreative opportunities, in some cases even infringing their procreative rights. It is less often considered how the procreative choices of the present generation can affect the procreative opportunities of later generations. In this paper we consider how procreation above the replacement fertility level now would limit the opportunities to procreate of future generations, potentially leading to inequalities in the opportunity to procreate between current and future generations. We suggest that procreative opportunity be treated as something akin to a finite extinguishable resource that can be distributed more or less equally across generations. We then consider how this bears on universal procreative rights. We suggest that if there is a universal right to produce some number of children, this number is no greater than the replacement fertility level for the global population. We then briefly discuss possible ways of enforcing demands of intergenerational justice in procreative opportunity.

¹Institute for Futures Studies (IFFS), Stockholm.

² Stockholm University Demography Unit (SUDA) & CEK (Centre for Cultural Evolution).

³ We are thankful to the participants at the workshop on Intergenerational Justice (UPF Barcelona, June 2019) and at the PPE seminar (IFFS Stockholm, April 2021) for helpful feedback, and to Serena Olsaretti and Erik Magnuson for helpful written comments. Financial support from Riksbankens Jubileumsfond (Sweden) (Grant No. M170372:1) is gratefully acknowledged.

Introduction

Reproductive decisions have in recent years received special attention. This is related to an increasing focus on how environmental constraints may limit the eventual size of the global population. The growing problem of climate change, for example, has prompted discussion about the extent to which an increased population can be a driver of climate change, as well as how to define the Earth's carrying capacity (Dasgupta, 2019, Greaves, 2018, Wynes & Nicholas, 2017). Environmental challenges may lead to a new set of constraints on the procreation opportunities of future generations, opportunities concerning both the number and the quality of life of the individuals who, under such circumstances, future generations will be able to bring into existence. These constraints on the procreation opportunities of future generations raise relevant questions of intergenerational justice.

Procreation rights are regarded by many as unconstrained and universal. External interference with the free exercise of these rights are rarely acceptable and the rights apply to all human beings irrespective of country, culture and context. Many hold the view that procreation rights are liberal with regard to the number of children that one may have. Michael Bayles, for example, claims "A human right to procreate involves an obligation on others not to limit a person's liberty to decide when and how many children he will have" suggesting the existence of a right of noninterference with one's decisions about the timing and number of one's children (1976, p. 42). Bayles' view is reflected in The United Nations' 1994 International Conference on Population and Development, which defines 'reproductive rights' as including "the basic right of all couples and individuals to decide, freely and responsibly, the number and spacing of their children" (ICPD, 1994: 13). The reference to all couples and individuals suggests that the right to decide the number of one's children is a universal basic right, one that applies to all individuals regardless of where and when they live. The World Health Organization (WHO), the main health organisation of the United Nations embraces this interpretation of procreation rights too. And the right of women to decide on the number and spacing of children is explicitly mentioned too in Articles 12 and 16 of the Convention on the Elimination of All Forms of Discrimination against Women, adopted in 1979 by the United Nations General Assembly, which obligates the party states to ensure "access to health care services, including those related to family planning" (UNGA, 1979).

The theoretical justification of this liberal interpretation of procreation rights is seldom spelled out. Sarah Conly has argued against a universal right to unlimited procreation on the grounds that no plausible theoretical basis can be provided for such a right. She points out that a right to determine the number of one's children can be justified based either on (i) important values for which procreation is necessary or (ii) a more fundamental human right, such as a right of non-interference with one's autonomous decision-making (Brock 2010, p. 382). According to Conly, neither (i) nor (ii) can establish a right to have as many children as one wants. The problem with appealing to (i) is that the values that procreation promotes (e.g. the meaning of life, personal identity, and dignity) can be secured by having one or two children (Conly 2005; 106-107). The problem with appealing to (ii) is that unlimited procreation could lead to a type of interference in the lives of others, frustrating interests which, plausibly, are more fundamental than interests in unlimited procreation (Conly, 107).

But let us suppose for the sake of argument that for many individuals there are important values that cannot be secured by having just one or two children, or, at least, that for many, there would be a significant gain in value from having a third (or fourth, or fifth) child. And suppose (perhaps implausibly) that it is possible for individuals to produce extremely large numbers of offspring without frustrating others' fundamental interests. Should we now recognize the existence of (more or less) unconstrained procreative rights?

We think not. In this paper, we argue that there are further reasons not to recognize a universal right to unlimited procreation. These reasons have to do with how individuals' procreative decisions affect the opportunities of future generations to make similar decisions. Specifically, we argue for the following constraint: if there is a universal right to create some number *n* of children, then *n* is no greater than the replacement fertility level for the global population--i.e., the average number of children born per woman at which the global population replaces itself across generations. Given the world's current demography, this implies, at the level of the individual, no more than two children per woman (The replacement fertility level is slightly above one child per person, or two children per woman.)

We show that if individuals did possess a right to have a number of children greater than the replacement fertility rate, then their rights would not be mutually exercisable. Moreover, the exercise of such a right by earlier generations would curtail the opportunities that later generations would have to procreate. Although the claim that there is no universal right to have more than two children does not imply that it is always impermissible to have that many, we suggest that having more than two children be treated as a privilege rather than as a right. While most philosophical arguments for limiting procreation (e.g. Ehrlich 1968; Conley 2009) are based on the assumption that more people will result in worse lives, environmental degradation, or other negative consequences, ours is concerned mainly with intergenerational inequality in the distribution of procreative opportunities.

The paper proceeds as follows. In Section 2, we show that given the physical

limitations of exponential population growth, if each individual possessed a right to create more than two children, then these individuals' rights would not be mutually exercisable. One potential consequence of this fact is that individuals' rights to have more than two children are non-compossible—if some individuals possessed the right then others would not. This would falsify the claim that the right in question is universal. In Section 3, we argue that given physical constraints on population growth, the exercise of a right to produce more than two children by earlier generations would lead to intergenerational inequality in procreative opportunity. We argue that this supports the claim that if there is a universal right to have some number of children, then this number is no greater than the replacement fertility level. In Section 4, we consider two further issues relating to intergenerational inequality of opportunities to procreate. We conclude in Section 5.

Rights to Procreate Above the Replacement Fertility Rate are not Mutually Exercisable

An important observation is that population growth above the replacement fertility level amounts to exponential population growth, which is unsustainable. (Inevitably, population size would reach a point at which population growth would turn negative.) That such growth is unsustainable is recognized in demographic theory, which involves, among other things, constructing models of the mutual relationships between population size and population growth. Classical examples include Malthusian models, or ecological models where available resources regulate population growth.

We assume there exists a population cap for which further human population growth is impossible. There is a sizable literature that attempts to estimate various possible population caps (see an exhaustive summary by Cohen, 1996) using all kinds of limiting factors (e.g., available habitable land, farming yield, or nitrogen for farming), finding very different estimates of what is the human global population cap. For our present purposes, we need only note that no matter what the population cap is--e.g. 10 billion, 50 billion, 500 billion, or higher--population growth above the replacement fertility level will eventually and quickly run up against the cap. We illustrate this in footnote 2 using demographic calculations⁴. That exponential population growth cannot continue indefinitely is an incontestable aspect

(1)
$$N_{t0} * (1+r)^t = N_t$$

⁴ Equation (1) below shows exponential population growth over time.

Population at time t is N_t and the starting population N_{t0} . The equation gives the eventual population size t years later given population growth r (r= 0.01 means a population growth of 1% a year).

of biological life. Hence, a fertility rate above the replacement fertility level is not possible over a sufficiently extended period of time.⁵

One upshot of the fact that exponential population growth is unsustainable is that if individuals did possess a right to produce more than two children (i.e. to have a number of children greater than the replacement fertility rate), then their rights would not be mutually exercisable (unless these rights are heavily qualified, a possibility we discuss below). If sufficiently many individuals were to exercise their right, others would be unable to do so. The point is not just that some who exist would be unable to have more than two children. That fact is trivial, guaranteed by the inevitable demise of the human race (e.g., due to the heat death of the universe) and the fact that those in humanity's final generation will have no progeny. The point is that if sufficiently many individuals had more than two children, this would

8*10⁹ * (1+0.01) ¹⁰⁰ = 21.6 * 10⁹

8*10⁹ * (1+0.01) ¹000 = 16.8 * 10⁹ * 10⁴

(2) $\ln(TFR/2)/\mu = r$

If we exemplify with a TFR of 3 (e.g., 1/3 of women have 2 children, 1/3 have 3 children, and 1/3 have 4 children and a mean age of childbearing of 30 (typically both for a country such as Sweden or Nigeria) we obtain:

ln(3/2)/30 = 0.0135

And if we use a TFR of 2.2 (e.g. 20 % of women have 3 children, and 80 % have 2 children) we obtain: $\ln(2.2/2)/30 = 0.00318$

In 2020, the global TFR was a little below 2.5

Below we show the implications of a TFR of 3 over 100 and 1000 years:

8*10⁹ * (1+0.0135) ¹⁰⁰ = 30.6 * 10⁹

8*10^9 * (1+0.0135) 1000 = 53.3 * 10^9 * 10^5

The calculations above illustrate the general point that any population growth quickly results in very high population growth, due to the very exponential character of growth. We therefore argue that within 20 or so human generations any constant substantive reproduction above replacement level fertility will therefore reach numbers that are likely completely unsustainable. If fertility is as high as in the 20 the century, this will happen within less than 5 generations.

⁵ On an extremely generous view of the future of human innovation and growth, it may be possible for humans to maintain modest population growth, over a large number of generations. But even in such a scenario, the limit to human population growth is only modestly higher than replacement level fertility (e.g., around 2.3 children per woman).

We use 8 billion for N_{t0} and show the effect of 1 % growth over 100 and 1000 years below (global population growth over the second half of the 20th century until today ranged between around 1 % growth to a little over 2 % in the 1960s). Clearly the population size given recently observed population growth becomes implausibly large.

We also relate different assumptions of the average number of births by women in a population. Population growth r for a given level of fertility (TFR) in a population, where μ is the mean age at birth or put differently, the average time between generations, in a population with very low mortality is given by equation 2 below:

For simplicity of calculation, we assume a very low mortality rate such as that of Japan 2020, where mortality has a trivial effect on population growth, and where TFR \approx NRR*2. For a country with mortality of a typical middle-income country such as Indonesia 2020, one would have to multiply the TFR by around 0.95 to get a comparable number.

remove other individuals' opportunity to have that many. Thus, in certain respects, the opportunity to procreate is similar to the opportunity to make use of a finite, exhaustible resource. If a sufficiently large number of individuals each use up more than some amount of the resource, then no one else will be able to use as much of it. The views of the World Health Organization and the United Nations stated in the Introduction portray a default conception of procreation rights as liberal or unconstrained. This default conception of procreation rights contrasts with our common understanding of other kinds of rights, including for example, rights to the use of natural resources. We do not normally think that there is a liberal or unconstrained right to the use and disposal of water, timber, or coal, that rightfully allows us to freely choose how much of these resources we may privately appropriate. Claims to these extinguishable resources from other currently existing and future individuals act as a constraint on how much of those resources each of us now can rightly appropriate. We suggest thinking of procreation rights in a similar manner.

The default conception of procreation rights as liberal/unconstrained is misguided, at least when put in the broader intergenerational context. This conception views procreation rights in the context of coexisting generations, where these rights can be infringed only by contemporaries, as when state authorities impose limits on procreation (think of China's one child policy, or the illegal sterilization of disabled individuals in different parts of the world).

However, when viewed in an intergenerational context, procreation rights seem to share important structural similarities with rights to use extinguishable resources. These similarities have traditionally gone unexplored. In this context, we suggest thinking of procreation opportunity as an extinguishable resource (like most natural resources) and to think of individual procreation rights as being constrained by the existence of a given finite number of procreative opportunities which, if appropriated in excess by some generations, would contribute to the impossibility of future generations satisfying their respective procreation rights, leading to a situation of intergenerational procreative inequality. On this view of procreation opportunities, other things being equal, the temporal placement of the different generations is a contingent criterion for how much procreation opportunities a given generation is entitled to.

What are the normative implications of the fact that individuals' rights to procreate at levels above the replacement fertility level are not mutually exercisable? The most important implication, we think, concerns intergenerational inequality. We discuss this implication in the subsequent section. However, there is at least one other potential implication that is worth mentioning here. If individuals' rights are not mutually exercisable, it may be that these rights are non-compossible--i.e. that they cannot co-exist. If individuals' rights (of a certain kind) are non-compossible, this spells trouble for the claim that the right in question is universal. However, whether the fact that rights are non-mutually exercisable entails that they are also non-compossible depends on further details about these rights.

Suppose that the right to have more than two children (a number greater than the replacement rate) is, or entails, a *claim right* to have more than two children. Assuming a Hohfeldian framework, the fact that one has such a claim right implies not only that it is permissible for one to perform whatever action falls under the right, but also that others have a duty not to interfere with one performing that action. Yet, given the existence of some population cap, some physical limit to population growth, if sufficiently many people were to have more than two children, this would make it physically impossible for others to do so. In this case, some people exercising a certain claim right would constitute *interference* with others' exercise of the same kind of right.

Steiner (1994) argues that in cases involving such interference, the individuals' claim rights are non-compossible. To illustrate, consider our earlier analogy involving finite exhaustible resources. For simplicity, suppose there are just two individuals, you and me, and a certain finite exhaustible resource, say a certain cake. Suppose that your eating more than half the cake precludes my eating more than half (and vice versa), and suppose for reductio that each of us has a claim right to eat more than half. But if you also have a claim right to eat more than half, then since my doing so interferes with your exercise of that right, my eating more than half violates a duty of non-interference, and hence, is impermissible. But an act cannot be both permissible and impermissible. It makes no sense to suppose that each of us has a claim right to eat more than half the cake non-compossible.

One response to this type of compossibility argument is that it rules out universal rights that we seem to have. Examples might be the right to freely choose one's profession or the right to freely choose whom to marry. The right to choose one's profession is constrained by, among other things, whether a sufficiently large number of people will choose to pursue the same profession at the same time, which can contribute to the collapse of demand for that particular type of profession, making it impossible for one to freely choose the profession one wants. Similarly, in monogamous marriages, people exercising their right to freely choose whom to marry (provided that the one they select also freely chooses to marry them) reduces the set of prospective partners whom others can marry.

We do not find such replies convincing, however. First, we don't think that individuals have *claim* rights of non-interference with choosing a career or spouse. The right to choose one's career or spouse is most plausibly construed as a liberty right. Moreover, there seem to be important differences between the exercisability of procreation rights, on one hand, and the exercisability of marriage or career rights, on the other. One difference concerns the different constraints that the exercise of these different kinds of rights impose on future generations. While single individuals can appropriate a large fraction of the marriage opportunities of a given generation (through e.g. poligamous practices), opportunities to marry are reset with every new generation. This imposes generational limits on the extent to which one can hoard marriage opportunities. Procreation rights differ from other rights in this respect.

A proponent of a universal right to have more than two children could respond to the charge of non-compossibility by denying that this right is, or entails, a claim right to unlimited offspring. Perhaps the right to have any number of offspring is merely a liberty right. In that case, it is not necessarily impermissible for others to interfere with one's having that many offspring. This would be a significant departure from tradition, however. Procreation rights are usually (and plausibly) construed as entailing some duty on the part of others not to interfere with those who wish to exercise them.⁶

Another possible response would be to claim that the right to produce more than two children needs to be heavily qualified or specified. For instance, one could claim that the right in question isn't simply the right to have more than a certain number of children, but more specifically the right to have any number of children *provided that* doing so would not deprive others of the opportunity to have as many, or increase the risk of significant resource depletion, or, ... etc. There are, however, familiar worries about such "specificationist" claims about rights. One worry is that the fully specified rights in question would be unknowable. Another is that they would lose most of their explanatory force.⁷

Yet another possible response would be to adopt a conception of rights that rejects compossibility as a requirement. Not everyone believes that the compossibility of a set of rights is a necessary condition of the truth of a moral theory that entails that set.⁸

Perhaps one of the abovementioned responses is plausible. (Whether this is so is a question beyond the scope of this paper.) But there are further reasons, of an egalitarian kind, to doubt the existence of a universal right to have more than n children. We turn now to our discussion of these reasons.

⁶ Consider, for instance, the following statement, which we take to be fairly mainstream:

The primary moral basis of reproductive freedom is individual or personal autonomy. Individuals' interest in autonomy is their interest in making significant decisions about their lives for themselves and according to their own values or conception of a good life, carrying out those choices *without interference from others*, and being free to revise their plans of life or conception of the good over time (Brock 2010, p. 86).

⁷ See, e.g., Thomson, 1990: 82-104 for a discussion of these concerns.

⁸ See e.g. Dowding and Hees (2004) for a critique of this claim.

Intergenerational Equality and Procreative Opportunity

Our main objection to the claim that there is a universal right to have more than two children is that the widespread exercise of such a right would produce intergenerational inequality with respect to procreative opportunities. For example, in some possible scenarios, if even a small fraction of the total number of individuals who ever live had more than two children, this would make it impossible for all individuals in later generations to do so if they wished.

In the context of intergenerational justice, it is common to appeal to some egalitarian principle of justice (Caney, 2016). The idea is that we have some duties of justice towards future generations and that those duties require an equitable distribution of some of the things that we think are of value.

The Lockean Proviso has been seen by some as a useful heuristic to think about distributive justice in the context of intergenerational justice (see Elliot, 1986; Arneson, 1991; Steiner, 1994; Wolf, 1995), and it is especially illustrative when thinking about finite resources. According to the Proviso, appropriation of a good is legitimate if there is enough and as good left in common for others (Locke 1980). Gosseries applies the proviso to the intergenerational context with the following formulation:

Each generation must leave to the next at least as much as what the next generation could have appropriated in the absence of any previous generation, or preferably, what the coming generation would otherwise have inherited if no previous generation had by its actions brought about a net improvement or a net deterioration. (Gosseries, 2008)

The proviso is highly underspecified to serve as a metric for the distribution of goods in the context of intergenerational justice. Among other things, it says little about what counts as "enough" and "as good as" for different kinds of goods (Gosseries, 2008).

As a heuristic though, the Lockean Proviso seems plausible, and it helps illustrate the distributive problem at hand. The exercise of an alleged universal right to procreate above the replacement fertility level by enough people in earlier generations, over the course of a given number of years, can make it impossible for any members of future generations to exercise the same right if they so wish. We find this to be a highly objectionable form of inequality with respect to individuals' opportunities to procreate. No matter how the Proviso is ultimately spelled out, it should at least yield the result that inequalities of this sort are unjust.

While welfare and resources are the most used metrics of intergenerational justice, opportunities for childrearing have recently been proposed as part of the

currency of intergenerational justice in the context of climate change and biodiversity degradation. This is the case of Gheaus' account on the right to parent (2016). Gheaus' proposal is sustained on the claim that people have a special interest in raising children in the right environment, to which climate change and environmental degradation are threats. According to Gheaus, this special interest grants a right to raise or parent children in adequate conditions that environmental degradation is a threat to. Therefore, she concludes, any generation that fails to pass on to the next generation a sustainable world is in fact preventing at least some of its contemporary children from making future voluntary choices with respect to childrearing opportunities (Gheaus, 2016).

Our proposal diverges from Gheaus' in that her proposal is concerned exclusively with the right to rear or to parent children in the right environmental conditions, while our proposal is concerned with procreative rights. Thus, Gheaus' account is insensitive to inequalities in procreation opportunities between generations, as long as people's parenting opportunities are shared equally with the members of future generations and that future generations also have a right to rear or to parent children in the right environmental conditions.

We agree with Gheaus that there is value in raising or parenting children (biological or not). Our argument here relies on a different but, we believe, more common value assumption; namely, that there is some special/added value in the organic unity of procreating together with raising one's own child. And that future generations will also have an interest in these two joint activities to the same extent that we do, other things being equal. It is the unequal distribution in the opportunity to procreate among different generations that this paper is concerned with.

Further Issues

In this section, we briefly touch on two further issues relating to intergenerational inequality of procreative opportunity, issues that we lack the space to fully explore in this paper but that we think are potentially fruitful areas of future research.

One issue concerns the relationship between inequality and responsibility for correcting unjust inequalities. Inequalities, including those that arise in the intergenerational context, are sometimes addressed by corrective principles of justice. For example, on Gheaus' account, responsibility for intergenerational inequality in the opportunity to rear children under the right environment is traceable by looking back at responsibility for climate change and environmental degradation. Thus, a principle of corrective justice along the lines of the Polluters Pay Principle (PPP) could be invoked to determine who is responsible for correcting the unequal distribution if possible, or for compensating those who are the victims of the inequality.

The example of intergenerational procreative inequality we have considered in this paper has a different source, however. In this case, causal responsibility for the inequality can be traced by looking at those who procreated above the replacement fertility level. If population growth above this level is bounded by a cap, then individuals' opportunities to procreate above the replacement fertility level will become more and more scarce as the population moves closer to the cap. Every generation procreating above that level will be causally responsible to some extent for the resulting inequality (perhaps they would be causally responsible to the same extent if the population growth rate was the same in each generation). We give an analytical and demographic example on how such reproductive opportunities would be distributed across generations in footnote 4.9 However, there are several reasons why it is hard to determine who is morally responsible for the violation of the future generation's right to procreation in non-idealized scenarios. Usually, the reproductive rate is divided differently among different members of the same generation for different reasons, be these cultural or economic, including excusable ignorance or need. Factors like these make it hard to invoke a principle of corrective justice like the PPP to be applied here (or, more exactly, a procreation version of that principle, what one might call the "Procreators Pay Principle").

(3)
$$N_t = KN_{t0}/(N_{t0} + (K - N_{t0}) * e^{-r_0 t})$$

⁹ Here we illustrate how excess reproduction above replacement fertility level will mean that given the inevitability of human population caps at some level, high reproduction of individuals today will have to be matched by lower reproduction of individuals in the future.

We can call the reproduction of a population of above a TFR from 2 "excess reproduction e" (current TFR=2+e) and the gap in population between current population size N_t and the size at which human population size is bounded N_{max}, "reproductive budget" b (N_{max} - N_t = b). Given constant fertility and some assumptions, b can be related to a value of TFR (i.e., e).

We can give all our assumptions above a mathematical form, by assuming that growth between $N_{\rm t}$ and $N_{\rm max}$ follows a certain path and we can use a logistic equation commonly used in ecology to model carrying capacity. This is a reasonable assumption, but any other function in which we will see monotonic decrease in fertility between the present and the time point at Nmax would give the same broad qualitative results. The logistic equation is therefore chosen to illustrate the broad qualitative patterns on how a reproductive budget would be distributed.

Mathematical form:

For a logistic growth equation, the relative population growth r decreases linearly over time until it reaches 0 for N_{max} or (K in most ecological models).

Population size at time t has an S-shaped form increasing rapidly in the beginning (where absolute growth is highest at the inflection point, N_t equals half of N_{max}).

The logistic equation is described in equation 3 below, where $r_{\rm 0}$ is the unconstrained population growth at t0.

Under such a constrained growth scenario the possible fertility of women will simply be given by the first figure, where TFR will be proportional to *r*. Clearly over time future generations of women will be forced to have lower and lower fertility until their fertility must exactly equal the replacement level fertility, where *e* must equal 0.

Another issue concerns avoiding unequal outcomes through policy. There are different ways to achieve balanced or sustainable population growth over generations that avoids the procreation inequality we have analyzed in this paper. Former policies aimed at containing population growth, like the Chinese One Child Policy have proven unpopular for their high degree of coerciveness in their application as well as their illiberal nature. Tradable procreation entitlements can be seen as an alternative policy aiming at containing population growth.

The risk of climate change and environmental degradation has attracted attention to the idea of a liberal individual climate budget (e.g., Caney, forthcoming). The idea is that every individual would be entitled to a certain amount of CO2 emissions, which would be divided between a range of different individual activities, including travelling, consumption, etc. The budget, liberally conceived, would allow for a substantial degree of freedom to the extent that each person could choose to engage in those activities that better satisfy her preferences as long as they are within the emissions budget, and eventually trade her remaining CO2 emissions in one activity domain for emissions on another domain. Reproduction could be seen as among the activities contained in a budget of this sort, for which a maximum amount of CO2 emissions would be allocated.

The argument spelled out here can serve as further support for a budget of this sort. At the same time, such a proposal would incorporate concern for other constraints on intergenerational parental inequality, like the environmental constraint that Gheaus' account is concerned with. How to specify a climate budget of this sort, including tradeoffs between the different emissions within the budget, as well as special allowances based on regional differences, medical conditions, etc. is a pressing challenge that we leave for future investigation.

Conclusion

Procreative rights have become a subject of increased scrutiny due in part to increasing anthropogenic climate change and the impact of increased population size on carbon emissions. There are reasons to recognize universal procreative rights in some form. But such rights need to be somehow constrained. We have argued for one specific constraint: if there is a universal right to have more than some number of children, this number is no greater than the replacement fertility level for the global population¹⁰. At the level of the individual, this entails no more than two

¹⁰ In our study we have argued for an absolute cap on global human population size. While we think this is not unreasonable, we note that if a person has a very optimistic view of human technological growth, and for example foresee human expansion beyond Earth, humans will still be bounded by population growth. For example, one could assume that technological growth and/or human expansion could

children per woman. If there were a universal right to have more than this, then given plausible demographic assumptions, individuals' rights would not be mutually exercisable. Most importantly, the exercise of the right by sufficiently many individuals in earlier generations would produce an objectionable form of intergenerational inequality in procreative opportunities. For this reason, we think, the replacement fertility level for the global population should function as a limit (though perhaps not the only limit) on the number of children that couples, and individuals, have a right to create.

References

Caney, S. 2014. "Climate change, intergenerational equity and the social discount rate". *Politics, Philosophy, and Economics.* https://doi.org/10.1177/1470594X14542566

Caney, Simon: "Liberal Egalitarian Emissions" Ms.

Cohen, J. E. 1996. *How many people can the earth support?* WW Norton & Company.

Dasgupta, P. 2019. *Time and the Generations: Population Ethics for a Diminishing Planet*. New York: Columbia University Press.

Dowding, K., & Hees, M. V. (2003). "The construction of rights." *American Political Science Review*, 97(2), 281–293.

Ehrlic, P. 1968. The Population Bomb. New York: Ballantine Books.

Gheaus, A. "The Right to Parent and Duties Concerning Future Generations", *Journal of Political Philosophy*, 24 (1):487–508.

Gosseries, A. 2008. "Theories of intergenerational justice: a synopsis", *Surveys & Perspectives integrating Environment and Society*, 1(1), http://journals.openedition.org/sapiens/165

sustain a population growth of 0.2 % (nearly) forever, even as that implies enormous population size after a few centuries.

All the arguments we put forward in this article would then still be applicable, but would instead relate to fertility levels corresponding to for example (as a hypothetical example) 2.3 children per women, and any claim to childbearing above that level (instead of replacement level) could be argued against on similar grounds as we do in this study. See also related discussion by Hardin (1959).

Greaves, H. 2018. "Climate Change and Optimum Population." The Monist.

Locke, John. 1980. Second treatise of government. C. B. Macpherson, Ed. Hackett Publishing.

Unesco, 1994. "International Conference of Population Development", *Programme* of Action:

https://www.unescap.org/sites/default/files/Programme_of_action_Web%20EN GLISH.pdf

Steiner, H. 1994. An Essay on Rights. Oxford: Blackwell.

Thomson, J. J. 1990. The Realm of Rights. Cambridge: Harvard University Press.

UNGA, 1979. "Convention on the Elimination of All Forms of Discrimination against Women", New York,

https://www.ohchr.org/en/professionalinterest/pages/cedaw.aspx

Wynes, K. & Nicholas. K. 2017. "The climate mitigation gap: Education and government recommendations miss the most effective individual actions." Environmental Research Letters 12(7): 074024.

Hardin, G. (1959). "Interstellar migration and the population problem." *Journal of Heredity* 50(2): 68–70.

H. Orri Stefánsson¹ What Is the Point of Offsetting?²

I shall defend two related claims. The first claim is that, contrary to what John Broome has argued, offsetting is not a legitimate way to meet one's duty not to cause greenhouse gas emission. The second claim is that, because of (the truth of) the first claim, we have no reason to offset rather than using the money to do more good in other ways. And since it turns out that offsetting is not an effective way to do good, we should not spend money on offsetting.

 $^{^1} Stockholm \ University \ and \ Swedish \ Collegium \ for \ Advanced \ Study, \ orristef ans son @philosophy.su.se, \ www.orristef ans son.is.$

² This paper was presented at the 5th Oxford Workshop on Global Priorities Research, where I received very helpful comments and suggestions. In addition, I have benefitted from discussing some of the ideas of this paper with Christian Barry. Special thanks to John Broome and Krister Bykvist, who both read the article and provided very useful written comments. Financial support from Riksbankens Jubileumsfond is gratefully acknowledged.

1. Introduction: Your emissions cause unjust harm

Hardly anyone denies that the climate is changing as a result of the greenhouse gas (GHG) emissions due to, for instance, our collective consumption and modes of transportation. Moreover, nobody can seriously doubt that these climate changes will harm many people and animals (although they may also benefit some people and animals). Some prominent philosophers however deny that the GHG emissions associated with a typical person's actions harm others (e.g., Sinnott-Armstrong 2005, Sandberg 2010, Kingston and Sinnott-Armstrong 2018, Budolfson 2019). In this paper, I shall assume that these 'individual denialists', as Broome (2019) calls them, are wrong. In particular, I shall take for granted that many of our individual actions—such as flying to Paris for the weekend or going on a Sunday ride in a gas guzzling SUV—unjustly harm others, due to their associated GHG emissions.

The estimates differ as to how much expected harm is caused by a typical person's emissions. And, of course, any such estimate will be highly uncertain. To take a few prominent examples, John Nolt suggests that the 'average American' causes, through their lifetime GHG emission, 'the serious suffering and/or deaths of two future people' (Nolt 2011: 9).³ John Broome (forthcoming), however, finds that the 'amount of killing' done by an average American's emission amounts to between 0.5 and 7 life-years. In earlier work, Broome pointed out that based on some prominent estimates of the social cost of carbon, 'the harm [due to GHG emission that] you do over a lifetime ranges between \$19,000 and \$65,000' (2012: 75). In that same work he also estimated the 'amount of killing' to be six months. Finally, a recent study finds that 'the lifetime emissions of 3.3 average Americans cause one excess death globally between 2020-2100' (Bressler, 2020: 1).

I do not know which of these estimates are correct, or most relevant, when evaluating the harm done by your emissions.⁴ In fact, the precise estimate does not matter much for my argument. So, I shall simply assume that the harm associated with your lifetime emission is the loss of 6 months of life. Now, these 6 months are very thinly spread. The assumption is not that your emissions will cause someone to die half a year earlier than they otherwise would. Rather, the assumption is that your lifetime emission is expected to cut lives short by six months *in total*, that is, when your emissions' effect on all people's lives is summed up.

³ Strictly speaking, Nolt only made the following *conditional* claim, without explicitly defending the antecedent: 'If over the next millennium as few as four billion people (about 4%) are harmed (that is, suffer and/or die) as a result of current and near-term global emissions, then the average American causes through his/her greenhouse gas emissions the serious suffering and/or deaths of two future people.' (Nolt 2011: 9)

 $^{^{\}rm 4}$ For the sake of simplicity, I shall use the phrase 'your emissions' to refer to the emissions that you cause.

Some might think that since the *expected* harm for each person is tiny, if the above assumption is correct, the expected harm in question is morally insignificant and can be ignored. I shall assume, without much argument, that that is not correct. Note that I need not assume that *imperceptible* harms are morally significant (nor that there are imperceptible harms).⁵ For even a shortening of a life by only a few minutes is perceptible. Nor need I assume that if only one person imposes a tiny expected harm, then that is morally significant. When it comes to the expected harms associated with your emissions, you are not alone. We who for instance travel and drive non-electric cars are all causing such harms. My assumption is that each of these expected harms is morally significant, but I shall remain agnostic about whether it would be morally significant if only one person inflicted such a tiny expected harm.

Moreover, I shall assume that these harms are *unjust*. In general, we have a justicebased duty not to harm others. The same holds, I contend, when it comes to imposing *risk* of harm on others. These duties can, of course, by overridden. For instance, we may sometimes by justified in harming another person who deserves punishment. But, following Broome (2012), I shall assume that this is not the case when it comes to the harm done by your GHG emission. Some of the people who suffer these harms do certainly not deserve them; for instance, people in developing countries who can themselves neither afford to fly nor drive. And these harms are often associated with actions that we do merely for our own pleasure (e.g., going for a Sunday drive or flying to Paris for the weekend), and these acts could be avoided at little cost to us. This is not true of *all* emission, however. But, to keep things simple, I shall not distinguish what some have called *luxury* emissions from *sustenance* emissions. Instead, I shall evaluate all GHG emission in terms of the harm it risks causing.

Still following Broome (e.g., 2012), I take it that we thus have a *justice-based duty* to have no carbon footprint. Justice demands of us that we do not harm others and that we do not risk harming others—except in the special circumstances that I assume do not hold when you emit. Therefore, justice demands that we do not emit. Such justice-based duties are owed to specific persons. In this they differ from what Broom calls 'duties of goodness', that is, general duties to do good. For instance, I owe you that I don't harm you. Correspondingly, you have a right not to be harmed. Therefore, I cannot satisfy this (justice-based) duty by harming you while preventing your friend from being harmed. In contrast, although I may have a duty to make the world better rather than worse, which I could satisfy by helping Syrian refugees, I do not owe it to any Syrian refugee that I help them, nor does each of them have a right that I help them. 'Your duty to have zero carbon footprint does not derive from

⁵ For discussion, see, e.g., Parfit (1987), Kagan (2011), Nefsky (2011), and Broome (2019).

your duty of goodness,' Broome (2012: 81) says. 'You must do it to avoid injustice'. As such, he argues, the duty in question is a duty owed to specific persons.⁶

So far, I have been in agreement with Broome. However, the main aim of this paper is to argue that Broome (2012, 2013) is wrong when it comes to his suggestion for how you can, at little cost to yourself, meet this duty to have zero carbon footprint, namely, by *offsetting*.⁷ I shall start by briefly describing GHG offsetting, and explain why Broome thinks that by offsetting you can perfectly meet your duty to not harm others through your GHG emission. I then argue that since the duty to have zero carbon footprint is grounded in a duty of justice, this duty cannot be met by emitting and offsetting. After that I briefly respond to a recent argument by Christian Barry and Garrett Cullity on offsetting.

My examination of Broome's and Barry and Cullity's arguments leads me to the conclusion that, first, we do not meet our duty to be carbon neutral by offsetting, and, second, that we have no reason to offset *rather than* giving to charities that do good more effectively than offsetting does. Thus, although we have *a* reason to offset, grounded in our general duty to do good, we have a *stronger* reason to instead give to a more effective charity.

2. Broome on Offsetting

When we offset the greenhouse gas emission caused by an activity, we pay someone to do things to ensure that the amount of GHG in the atmosphere remains the same as what it would have been had we not engaged in the activity in question. Several companies offer this type of service, and promise to offset everything from an individual flight to a whole life. In theory we could do the offsetting ourselves, rather than paying someone else to do it (e.g., by planting lots of trees), but in practice that is infeasible for most of us.

There are two types of GHG offsetting. One type, which Barry and Cullity (ms.) call o*ffset by sequestering*, consists in removing GHG molecules from the atmosphere. The most common way to do so is by planting trees. It is possible to mechanically remove carbon dioxide from the atmosphere and turn it into rocks, but that is

⁶ Actually, Broome (2013: 6) also seems to suggest that the duty to not harm people is a duty owed to specific persons *whether it is a duty of justice or not*. That would mean that it is not really important for my argument whether we classify the duty not to harm (and the duty not to risk harming) as a duty of justice or as a duty of goodness. What matters, for my purposes, is that the duty not to harm is a duty of owed to specific persons. Nevertheless, I shall continue to assume that the duty in question is a duty of justice, which implies that it is a duty owed to specific persons; but strictly speaking I only need to make the weaker and implied assumption.

⁷ Note that I will only be considering offsetting by individuals, not by companies or countries, and I will not discuss the related practice of engaging in a cap-and-trade system.

still very expensive and too water-intensive to be sustainable in most parts of the world. The other type, which Barry and Cullity (ms.) call offset by forestalling, consists in preventing others from emitting GHG. Examples include funding green energy and providing homes in developing countries with energy efficient cooking stoves (see, e.g., Broome 2012: 85-88).

Offsetting is currently very cheap. There are two main (related) reasons for this (Broome 2012, Spiekermann 2014). On the one hand, many of the projects that offsetting funds take place in developing countries, where for instance labour and land is cheaper than in the developed world. On the other hand, and relatedly, there is still very little demand for offsetting, which keeps the price low. If more people chose to offset, then it would no longer be possible to meet the demand with only cheap offsetting projects, e.g., in developing countries, and the price would rise. But given the current low demand one can, for instance, offset a return flight between London Heathrow and Stockholm Arlanda for as little as £13.

It is important to keep in mind that when one offsets one's emissions, one either pays for offsetting that has already taken place, or one funds offsetting that will take place in the future. For instance, one may pay for, say, completed green energy projects that were funded by loans, or one may pay for trees to be planted in the future. Therefore, when one offsets a Sunday drive, say, there will inevitable be some time (in particular, immediately after the drive) during which there is more GHG in the atmosphere than there would have been had one not gone for the drive (keeping everything else fixed). Nevertheless, by offsetting, the hope is that in the long run, or on balance, you cause no more GHG to be added to the atmosphere than had you not gone for the Sunday drive. The issue is somewhat more complicated when it comes to financing already started (perhaps even completed) offsetting projects (that were, say, originally financed by loans). But the hope is that by doing so, you contribute to the continuation of the offsetting market, and encourage companies to continue their offsetting. So, by offsetting all your emissions, the hope is that you cause no more GHG to be added to the atmosphere than had you never existed; thus, you cause zero net GHG emission (and achieve 'carbon neutrality').

Broome thinks that by offsetting all your emissions, you satisfy your justicebased duty to be carbon neutral: 'If you successfully offset all your emissions, you do no harm by emissions. You therefore do no injustice by them.' (2012: 85) He also claims that by doing so, your emissions do not harm *anyone*: 'If you offset all your emissions [...], you make sure that your presence in the world causes no greenhouse gas to be added to the atmosphere. You therefore do no harm to anyone through your emissions.' (2012: 87)

In the next section I shall argue that Broome is wrong. Even if you offset all your emissions, you do—or, at the very least, you have reason to believe that you do—

harm some people, without compensating them, through your emission. Therefore, I argue, by emitting and offsetting you do injustice to these people (and you do not satisfy your duty of justice).

3. Against Broome on Offsetting

As discussed in detail by Morgan-Knapp and Goodman (2015)—and, in fact, recently discussed Broome too (in his 2019)—even comparatively small differences in the concentration of greenhouse gas can have huge impact. One reason is that the climate is 'chaotic', which implies that tiny interventions can have huge impact. A related reason is that there are 'meteorological thresholds', such that when GHG concentration, or the corresponding temperature, passes a threshold it causes a storm, or a flood, or a drought, or some other potentially harmful event.

Moreover, as Morgan-Knapp and Goodman (2015) point out, the risk from, say, going on a Sunday drive in a gas guzzling SUV is *asymmetric*, in the sense that it is more likely to cause a climate-related harm than it is to prevent such a harm. Increased GHG concentration is correlated (at least given the current and relevantly close levels of concentration) with extreme weather events, such as storms, floods, and droughts. Therefore, although we cannot know whether a particular emission will cause such an event, or instead prevent such an event, or have no effect on any such event at all, we do have reason to believe that the act is more likely to cause such an event than to prevent it. Acts that emit GHG are thus different from acts such as going for a walk, which in theory could cause an extreme weather event, but which are just as likely to prevent such an event.⁸

Furthermore, recall that by offsetting an individual action, one is not preventing emission from *that* action. By emitting and offsetting you in fact can be almost certain that there is a period (in particular, immediately following the emission) at

⁸ However, even seemingly innocent acts such as going for a walk are very likely to cause *some* harm *in the very long run*, or so at least MacAskill and Mogensen (2019) argue. One might view MacAskill and Mogensen's argument as a *reductio* of the idea that we have a justice-based duty not to harm others. After all, if there is such a duty, then it would seem to follow from their argument that we have a duty to as little as possible in our lives. However, one can also view their argument as supporting my conclusion, by vindicating a more general claim from which my conclusion follows. If MacAskill and Mogensen's argument is sound, then it is in general pretty much impossible to satisfy our justice-based duty not to harm others. The only way to satisfy it is to do 'nothing'. But if that is the case, then it would seem to follow that we cannot satisfy our justice-based duty not to harm others by emitting and then offsetting we do something. Therefore, one can view MacAskill and Mogensen's argument as supporting my main claim, by vindicating a stronger claim. However, the converse is not true; MacAskill and Mogensen's general claim—that is, that the only way in which one could possible avoid harming others, in the long run, is by doing nothing—does not follow from my more modest claim—that is, that by offsetting we do not satisfy our duty not to harm others. Therefore, even those who do not accept MacAskill and Mogensen's argument might be sympathetic to mine.

which there is greater concentration of GHG in the atmosphere than there would have been had you not emitted. (Note that this is true even if the offsetting takes place *before* your emission. The difference with offsetting that takes place after the emission is merely that, in the case where the offsetting takes place prior to the emission, there is less concentration of GHG *before you emit* than there would have been had the offset not taken place. But that does not change the fact that had you decided not to emit, despite the offset, then there would have been less GHG in the atmosphere.) If the offsetting is successful, it nevertheless means that in the long run your being in the world leaves the concentration of GHG unaffected.

Moreover, since your emission and offsetting effects the concentration of GHG in the atmosphere, even if only periodically, it changes a state of the atmosphere that, due to its chaotic nature, will almost certainly affect the weather in some way; it will cause some extreme whether events and prevent others. This in turn will affect the occurrence or not of later extreme weather events. As a result, the pattern of such events over, say, the next hundred years will be different from what it would have been without your emission and offsetting. Therefore, the distribution of *harm* caused by such events will also be different. So, some people will be harmed who would not have been harmed had you not emitted and offset.⁹

Another way to put the above point is that when you emit and offset it is the case that (and you moreover have reason to believe that):

- There is an event *E* (e.g., a storm, flood, drought, ...), which seriously harms someone if it occurs, and a probability p>0 that your emission causes event *E*.
- There is *zero* probability that your offsetting prevents event *E*: *if* your emission-without-offset causes event *E*, then your emission-and-offset also causes event *E*.
- There is however some probability q>0 that your offsetting prevents some event *F* (e.g., a storm, flood, drought, ...), which seriously harms someone if it occurs.
- The people who would be harmed by *E* are not those who would be harmed by *F*.

Here is yet is another way to put the above point.¹⁰ Recall that there are multiple meteorological thresholds, such that when concentration of GHG (or the corres-

⁹I am grateful to John Broome for having very graciously helped me improve my criticism of his argument by suggesting the formulation of my criticism in this paragraph (or something close to it).

¹⁰ I have benefited from discussing the argument of this paragraph with Hilary Greaves and Charlotte Unruh (at the 5th Oxford Workshop on Global Priorities Research).

ponding temperature) passes such a threshold, a potentially harmful event occurs, such as a drought, storm, flood, etc. When you emit and offset, you have reason to believe that your emission may cause one threshold to be passed (sooner than it otherwise would have), which leads to some event E. You also have reason to believe that your offsetting may prevent another threshold from being passed (or delay it being passed), which would have led to some other event F. Events E and F may be quite different, and they would for sure take place at different times. Therefore, the people who would be harmed by E are most likely not the ones who would be harmed by F.

Now, recall that I have been assuming that your duty to have zero carbon footprint is grounded in the justice-based duty not to harm others. As such, it is a duty owed to specific people: the people who would be harmed by your emission. In fact, it is a duty owed to *each* person. Correspondingly, each person has a right not be harmed. Since a duty not to harm is a duty owed to specific persons, I cannot—except in special circumstances—justify harming one person by preventing other people from being harmed; each person has a right not be harmed, which cannot be violated *merely* for the sake of preventing another person from being harmed.

In light of the above, I find it hard to see how emitting and offsetting could be a way of satisfying one's duty not to emit GHG.¹¹ By emitting and offsetting you have reason to believe that you harm some person while preventing someone else from suffering harm. Those who are harmed by the emission that we offset have a *right* not to be harmed, that cannot be violated *merely* on the grounds that one will prevent someone else from suffering a similar harm. In fact, since those who may be harmed by one's emissions (typically) have just as strong claim not to be harmed as those who would be harmed had it not been for one's offsetting, it is hard even to see how one's conduct can be just overall when one emits and offsets. In sum, *if* the duty to be carbon neutral is grounded in a justice-based a duty not to harm, that is, grounded in a duty owed to each person—as I have been assuming, following Broome—then it is not true that we can satisfy this duty by offsetting our emissions, contrary to what Broome suggests.

Moreover, it is false that you 'do no harm to anyone through your emissions' if you offset them all, as Broome (2012: 87) claims. The *aggregate harm* due to GHG emission in the long run *may* be no greater if you offset all your emissions than it would have been had you not emitted (and not offset) at all. But that does not, of course, mean that your emissions do not harm *anyone*. In fact, when you emit and offset, you most likely harm someone while preventing someone else from experiencing harm.

¹¹ Torpman (2014: 194) comes to a similar conclusion.

4. Barry and Cullity on Offsetting

Barry and Cullity (ms.) argue that by offsetting all our emissions, we do not impose hard-to-justify risks on others. They do admit that emitting and offsetting is 'not a way of ensuring that my emissions make no difference to who is harmed'. Moreover, they admit that even if you offset all your emissions, there will be time-periods when the atmospheric concentration of GHG is higher than what it would have been without your emissions. Nevertheless, they think that: 'By paying for offsets, I can act in a way that carries the expectation of leaving the atmosphere with no greater concentration of GHG than if I had emitted nothing. If so, I expose climatevulnerable people to no additional risk.' (27)

Does the second sentence follow from the first? Suppose that by offsetting 'I can act in a way that carries the expectation of leaving the atmosphere with no greater concentration of GHG than if I had emitted nothing'. Does it follow that 'I expose climate-vulnerable people to no additional risk'?

The second sentence does follow from the first if understood as the claim that my lifetime behaviour causes no net increase in the risk to climate-vulnerable people *as a time-extended population*. In other words, take the set of all climate-vulnerable people who will ever exist, from today onwards. If I successfully offset all my behaviour, then my being in the world does not expose this population to any more risk than had I never existed.

However, the second sentence does *not* follow from the first if understood as a claim about the harm—nor in fact if understood as a claim about the risk of harm—inflicted on *specific* individuals. Recall, from the last section, that when I emit and offset, I have reason to believe that I will harm some people (by causing event E), but, in terms of the total harm from GHG emission, I may leave things as they would have been had I not emitted, since I prevent other people from experiencing harm (by preventing event F).

Since the duty to be carbon neutral is a duty of justice, the latter interpretation of the second sentence would have to be true for offsetting to satisfy that duty. In other words, since duties of justice are owed to specific persons, you could only fully satisfy your duty of justice by emitting and offsetting if you had reason to believe that in doing so there is no person whom your emission will harm. But that is false, as we have seen. Hence, I contend, you don't satisfy this duty by emitting and offsetting.

Barry and Cullity however point out that the magnitude of risk during the timelag between your emission and your offsetting will be very small: If I offset my emissions completely by the end of my life, the expected harm associated with a time-lapse between emitting and offsetting [...] will be relatively small—comparable to the many other small risks created by much of our everyday activity, and therefore similarly easy to justify. (29)

This may be true. But wouldn't it also be true also of emission *without* offsetting? Apparently, lifetime odds of dying in a car accident in the US is about 1/100.¹² As far as I know, there is no available data on the lifetime odds that a driver *kills* someone when driving. But it is not unreasonable to assume, I think, that it is similar to the lifetime odds of dying in a car accident. Now, suppose that the average numbers of years lost when non-drivers die in car accidents is 50. I am not sure if this is true, but it could be true if young people are more likely than old people to be killed in car accidents where they are not the driver.

Given the above assumptions, the expected harm of a lifetime of emitting *without offsetting* (according to, for instance, Broome 2012) is comparable to the expected harm (to others) of driving. Therefore, it would seem that by Barry and Cullity's logic, emitting without offsetting is similarly easy—*or hard*—to justify as driving. What should we conclude from this? Are we justified in emitting without offsetting? Or should we somehow offset the expected harm from our (electric-car) driving?

There might be some relevant difference between (electric-car) driving and emitting, which could explain why we are required to offset our emission but not the risk we impose on others by driving electric cars.¹³ Perhaps we all have reason, *ex ante*, to agree to a system that allows driving, and maybe the costs and benefits—and the risks—from such a system are largely reciprocal. The same does not seem to be true of emitting without offsetting. In particular, future generations would have a reasonable complaint against such a system, and they would bear more of the cost and less of the benefit than current and past generations.

However, I think that all this is a bit of a red herring. It may be that the risk of harm associated with a time-lapse between emitting and offsetting is sufficiently small that we accept it as a necessary evil, given the system with which we are stuck and given our other goals. But that does not mean that when you emit and offset you impose no additional risk on *any* climate-vulnerable people. In fact, in the last section we saw that, when you emit and offset, you do impose additional risk on *some* people (during some times and in some places) while reducing the risk on *other* people (at other times and in other places). Moreover, as we also saw in the last section, you have reason to believe that while your offsetting prevents someone

¹²See, e.g., https://www.nytimes.com/2019/01/14/us/opioids-car-crash-guns.html.

¹³ Thanks to Christian Barry for a very helpful correspondence about this issue.

from being harmed, your emission *actually* harms someone else (as opposed to 'merely' *risking* harming someone else).

In sum, we have not yet seen an argument that by offsetting all your behaviour you 'do no harm to anyone through your emissions', as Broome claims, nor that you then 'expose climate-vulnerable people to no additional risk', as Barry and Cullity claim. In fact, even though you offset all your emission, you can be pretty much *sure* that you inflict harm on *some* climate-vulnerable people, and you can be *certain* you impose additional *risk* of harm on them. So, you do not satisfy your duty to not harm these people.

5. Conclusion: Give to More Effective Charities Instead

I have argued that by emitting and offsetting, we do not meet the justice-based duty that we owe to each person to not harm them. Does this mean that we have *no* moral reason to offset our emissions? Not necessarily. Recall that, in addition to having duties of justice, we also have duties of goodness that are not owed to anyone in particular. It is possible, then, that we have a moral reason to offset, stemming from our moral duty to do good.

The problem, however, is that apparently, we can do more good by giving to effective charities. For instance, Broom (2013: 9) himself points out that 'if you aim to use your resources to improve the world, reducing emissions of greenhouse gas is not the way to do it. To improve the world, you should carry on emitting, and send the money you save by doing so to a tuberculosis charity.' (See also Broome 2012: 66, 81.)

So, if the *only* reason for offsetting is to do good, then we have a *stronger* reason to give to an effective charity than to offset. I do indeed believe that the only reason we have for offsetting is to do good. We have a duty (of goodness) to do good, and one way of doing good is by offsetting. And that is the only reason we have for offsetting, I believe, since I take my above argument to establish that offsetting is not a way of meeting our justice-based duty not to harm others. But offsetting is not an effective way to do good. Therefore, if you are considering offsetting your greenhouse gas emissions, you should give the money to a more effective charity instead.

Bibliography

Barry, Christian, and Cullity, Garret. ms. Offsetting and risk-imposition. Unpublished manuscript.

Bressler, R. Daniel. 2020. The mortality cost of carbon. *CEEP Working Paper Series* 11.

Broome, John. 2012. *Climate Matters: Ethics in a Warming World*. W. W. Norton & Company.

----- 2013. The Public and Private Morality of Climate Change. *The Tanner Lectures on Human Values* 32, 3-20.

----- 2019. Against Denialism. *The Monist* 102, 110-129.

----- 2021. How much harm does each of us do? *Philosophy and Climate Change*. Ed. Mark Budolfson, David Plunkett, and Tristram McPherson. Oxford: Oxford University Press.

Budolfson, Mark. 2019. The inefficacy objection to consequentialism and the problem with the expected consequences response. *Philosophical Studies* 176, 1711–1724.

Kagan, Shelly. 2011. Do I Make a Difference? *Philosophy and Public Affairs* 39, 105–141.

Kingston, Ewan, og Sinnott-Armstrong, Walter. 2018 Kingston. What's Wrong with Joyguzzling? *Ethical Theory and Moral Practice* 21, 169–186.

MacAskill, Willilam, and Mogensen, Andreas. 2019. The paralysis argument. *GPI Working Paper* 6.

Morgan-Knapp, Christopher and Goodman, Charles. 2015. Consequentialism, Climate Harm and Individual Obligations. *Ethical Theory and Moral Practice* 18, 177–190.

Nolt, John. 2011. How Harmful Are the Average American's Greenhouse Gas Emissions? *Ethics, Policy and Environment* 14, 3–10.

Nefsky, Julia. (2011), Consequentialism and the Problem of Collective Harm: A Reply to Kagan. *Philosophy & Public Affairs* 39, 364–395.

Parfit, Derek. 1987. Reasons and Persons. Oxford University Press.

Sandberg, Joakim. 2010. My Emissions Make No Difference. *Environmental Ethics* 33, 229–248.

Sinnott-Armstrong, Walter. 2005. It's Not My Fault: Global Warming and Individual Moral Obligations. *Perspectives on Climate Change*. Ed. Walter Sinnott-Armstrong and Richard Howarth. Elsevier.

Spiekermann, Kai. 2014. Buying Low, Flying High: Carbon Offsets and Partial Compliance. *Political Studies* 62, 913–929.

Torpman, Olle. 2014. Review of John Broome *Climate Matters – Ethics in a Warming World. Theoria* 80, 191–195.

Göran Duus-Otterström¹

The Role of Subsistence Emissions in Climate Justice²

The climate justice literature typically endorses a moral right to emit subsistence emissions, that is, emissions that are necessary adequately to satisfy vital interests. Yet given the urgency of reducing greenhouse gas emissions, endorsing this right can seem tantamount to a moral permission to cause dangerous climate change, suggesting that it must be tempered. This paper argues, however, that there is no reason to think that a moral permission to produce subsistence emissions entails an exemption from remedial responsibility. Recognizing the right to produce subsistence emissions is compatible with avoiding very significant climate change if many people could compensate for their subsistence emissions without jeopardizing their vital interests.

¹University of Gothenburg and Institute for Futures Studies, goran.duus-otterstrom@iffs.se.

² I am grateful to Andreas Bengtson, Bengt Brülde, and Henrik Friberg-Fernros for providing written comments on their working paper. However, this version of the paper does not benefit from their many valuable comments as I did not have time to implement them. Nevertheless I want to thank them prospectively. Financial support from Riksbankens Jubileumsfond (grant M17-0372:1) is gratefully acknowledged.

1. Introduction

If humanity is to avoid (even more) dangerous climate change, total greenhouse gas (GHG) emissions must be capped. This raises an important ethical question: How should the remaining emissions be allocated? Ever since the publication of Henry Shue's (1993) paper "Subsistence Emissions and Luxury Emissions", the concept of *subsistence emissions* has played an important role in academic discussions about how to answer question. Shue introduced³ this concept in criticizing the idea that emissions should be cut where the economic costs would be the lowest. Shue's argument was that such an approach, while economically efficient, would be grossly unjust since it would ignore "the fact that some sources [of emissions] are essential and even urgent for the fulfillment of vital needs and other sources are inessential or even frivolous" (ibid, 55). If we are to allocate whatever is left of the carbon budget in a just way, he maintained, we must recognize such "qualitative" differences between emissions, and give strong if not absolute priority to emissions that are essential for "survival or decency" (ibid, 55).

Shue's message resonated, and continues to resonate, deeply in subsequent theorizing of climate justice. Yet the concept of subsistence emissions is a bit like conjunctions in natural languages: often used but rarely analysed—and surprisingly puzzling once one investigates them more closely. I want to shed some light on subsistence emissions, both conceptually and normatively. I address two questions. First, what are subsistence emissions? How should we define them? Second, what is the normative force of subsistence emissions? How does the fact that a set of emissions are of the subsistence variety matter for whether someone has a right to produce them?

My argument with regard to the first question is that contrary to what some commentators suggest, the concept of subsistence emissions is neither vague nor ambiguous. Emissions are of the subsistence variety whenever they are necessary adequately to satisfy a vital interest, nothing more and nothing less. It is true that what counts as a "vital interest", or a satisfier of such an interest, is not always clear; this might vary with, for example, the level of economic development of a community. But all this shows is that the concept of subsistence emissions depends on a prior account of our vital interests. The concept itself, I shall argue, is precise.

With regard to the second question, however, things are less straightforward. Some theorists equate subsistence emissions with an exemption from mitigation

³ As Shue (2019, 251-52) notes, he was drawing on earlier work by Agarwal and Narain (1991), but it is fair to say that he made the concept of subsistence emissions famous. Shue also drew on his seminal book *Basic Rights*, in which Shue argued that subsistence rights—the socially guaranteed enjoyment of "unpolluted air, unpolluted water, adequate food, adequate clothing, adequate shelter, and minimal preventive public health care"—are a precondition for enjoying other rights (2020, 23).

burdens. The thought is that no one should be held cost responsible for having emitted subsistence emissions. But this overlooks that while no one can be reasonably expected to refrain from *producing* subsistence emissions, this does not rule out that people have a moral duty to *compensate* others for these emissions afterwards, in the form of emissions reductions or in other ways. Drawing on the distinction between emitting with and without compensation, I argue that subsistence emitting only furnishes an exemption from mitigation burdens for people whose vital interests would be frustrated, or frustrated further, if they were to compensate. This makes subsistence emissions less morally forceful than standardly thought, and serves to blunt their potentially radical implications.

The paper is structured as follows: in section 2, I offer a conceptual analysis of subsistence emissions and note some important features of the proper definition of this concept. In section 3, I proceed to discuss subsistence emissions from a normative perspective. More specifically, I argue that what I call the "exemption claim"— the idea that emitters should not be expected to take climate mitigation burdens because of emitting subsistence emissions—should be rejected since some subsistence emitters may only have a "qualified" moral right to emit. I also draw out the implications of this argument for climate justice. In section 4, I discuss some objections. Section 5 concludes.

2. Subsistence emissions

Let us begin by looking at how the concept of subsistence emissions has been described by prominent climate ethicists. Shue understands them as emissions that are "essential … for either survival or decency" (Shue 1993, 55). Simon Caney defines them as "emissions required for [persons] to attain a minimal decent standard of living" (Caney 2009, 138). Steve Vanderheiden defines them as "a level of emissions sufficient to allow for … basic human functioning" (Vanderheiden 2008, 243).⁴ Finally, opting for a somewhat eclectic approach, Dominic Roser and Christian Seidel define them as "the emissions required for survival, for a minimally decent life, for meeting the most important basic human needs, and to ensure that human dignity and human rights are respected" (Roser and Seidel 2017, 144). This is disparate and potentially conflicting list of accounts, but the core message is clear: emissions only count as subsistence emissions if they produce the things a person needs to reach some basic moral minimum—"the line beneath which no one is to be allowed to sink", to use Shue's memorable phrase (Shue 2020, 18). I will capture this

⁴ To be precise, Vanderheiden (2008) uses the term "survival emissions", but this is because he adopts a conservative theory of vital interests, as I explain shortly.

by saying that subsistence emissions satisfy (protect or promote) a *vital interest.*⁵ It is important to note, however, that satisfying vital interests is not enough, for that would mean that any emissions that go into, say, producing food are subsistence emissions regardless of whether we are talking about a 100 dollar steak or a bowl of rice. For emissions to count as subsistence emissions, it must also be the case that they are *necessary* to satisfy vital interests. Emissions that satisfy a vital interest would not be subsistence emissions if the interest in question could be satisfied without emitting anything at all, for example. This is captured by the accounts above, which stress that the emissions must be "essential" or "required" for reaching, say, a basic level of functioning.

Putting these thoughts together, I propose the following definition: a set of emissions counts as "subsistence" emissions if and only if:

- 1. The emissions satisfy a vital interest.
- 2. At the time of emitting, there is no reasonable alternative way of satisfying this vital interest to an adequate degree.

The components are individually necessary and jointly sufficient for emissions to count as subsistence emissions. The reason to add the qualifier "to an adequate degree" is that vital interests only turn emissions into subsistence emissions when they are not satisfied.⁶ Food is a vital interest, for example, but that does not make the emissions associated with gluttonous overeating subsistence emissions. Only emissions necessary for being *adequately* nourished count as subsistence emissions. Whether emissions are of the subsistence kind also depends on whether it is reasonable to expect people to avoid them, more on which below and in section 3.

It is sometimes claimed that subsistence emissions is not a very precise concept. Roser and Seidel, for example, note that while no one doubts that "the emissions of a bus ride ... undertaken by a person in extreme poverty in search of work" are subsistence emissions, things are much less clear when it comes to "an intercontinental flight to celebrate a brother's wedding" (Roser & Seidel 2017, 176). They conclude that "boundary between subsistence emissions and luxury emissions is less clear than it may appear at first sight" (ibid.). But if there is any unclarity here, this is not

⁵ A possibly more natural way to put it is that subsistence emissions satisfy a *basic need*, but appeals to basic needs are subject to controversy, especially because they are rooted in the idea of a universal human nature and invite questions about paternalism (see e.g. Crocker 2008, 80–81, 129–41). For this reason, I will link subsistence emissions to vital interests, the idea being that this term may be less loaded. The difference is largely semantic, though, since both "vital interests" and "basic needs" refer to, in Griffin's famous words, "what we need to survive, to be healthy, to avoid harm, to function properly" (Griffin 1986, 42; see also Shue 2020, 126).

⁶ On this point, see Griffin (1986, 51).

because the *concept* is unclear but because it is unclear which interests are vital in the relevant respect. I am not suggesting that Roser and Seidel overlook this point, but the risk is that criticism such as theirs comes across as saying that the notion of subsistence emissions itself is vague or ambiguous when in fact it is neither. Emissions are subsistence emissions if and only if they are necessary to satisfy a vital interest to an adequate degree, and so if close family relations is a vital interest, and maintaining close family relations requires that we go to our brother's wedding, then the emissions from taking an intercontinental flight to our brother's wedding just *are* subsistence emissions, absent other reasonable means of transportation.

Some might resist the thought that something as inessential as flying to a wedding could count as "subsistence emissions". It can be tempting to think that subsistence emissions should be strictly about "those material provisions required for enjoying a minimal physical and physiological well-being" (Mancilla 2019, 2). But this just underlines that the concept of subsistence emissions depends on a prior theory of vital interests. Some such theories would indeed reject that attending our brother's wedding could count as a vital interest.⁷ This is especially true for theories that restrict vital interests to what it takes to live a life of average length without serious physical impairment. Yet there are other theories that come to a different conclusion. Many philosophers argue that people have a vital interest in leading a minimally "decent" life, where this includes things like having the capacity for autonomous agency and social participation.⁸ David Miller, for example, suggests that there is a human need not to be degraded in or excluded from one's society (Miller 2007, 181). Such theories may well find that flying to our brother's wedding produces subsistence emissions; it depends on whether the opportunity for close family relations is part of a minimally decent life. Note that the two types of theory are not mutually exclusive. Since there is no decent life without life, decency-based theories cover everything survival-based theories cover, but they add additional items to the list. Thus, survival-based theories pick out a proper subset of decency-based theories.

Part of the reason the concept of subsistence emissions can seem vague or ambiguous is that what counts as vital interests appears culturally and socioeconomically variable, especially if we opt for a decency-based account. As Christian Baatz (2014) has noted, once a society develops in a carbon-intensive way, it may "lock in" ways of life which require extensive GHG emissions. For example, a society

⁷ At least typically. Any act could be derivatively connected to a survival interest. If your brother is the only person capable of curing your life-threatening illness, and he will not treat you unless you attend his wedding, then flying to his wedding would produce subsistence emissions even on a survival-based account.

⁸ For valuable discussion, see Brock and Miller (2019). Baatz (2014) explicitly defines the ends of subsistence emissions in terms of decent lives.

that has planned its infrastructure around private motoring may well generate a vital interest in owning a car. This may seem to make the concept of subsistence emissions too flexible—anything could be a source of subsistence emissions if the activity producing the emissions is seen as important enough. However, it is important to distinguish here between vital interests and the "satisfiers" of such interests (Gough 2015). It does seem true that what counts as a decent life varies over time and between societies, but this is not because the vital interests are different as much as it is because the means of *satisfying* those interests are different. So while it may initially seem puzzling that owning a car could count as a source of subsistence emissions in one context but not in another, this impression goes away once we remember that driving a car merely happens to serve a vital interest which in turn is not variable (e.g., to have adequate means of transportation in one's society). It is not strange that the necessity of owning a car differs depending on whether we live in Montana or Manhattan.

Let me end this section by stressing two further points. First, the popular distinction between subsistence emissions and luxury emissions really is too crude. Just because emissions are not necessary adequately to satisfy vital interests, that does not mean that they serve sheer pleasure, as the term "luxury" suggests. For example, though neither is a source of subsistence emissions, there is a recognizable moral difference between going for a joyride in one's SUV and using it to pick up one's kid at kindergarten when one is too lazy to take the bike. The relevant distinction should be between subsistence emissions and non-subsistence emissions, where the latter comprises a ramp of emissions sources of increasing frivolity. Second and more importantly, subsistence emissions are not exclusive to the poor. This point is easy to overlook since the concept of subsistence emissions is intimately linked to the need for economic development. Indeed, Shue's concern in defending the concept was precisely to point out that economic development bestowed a different weight to the emissions of the poor (Shue 1993, 58). Nevertheless, under our current energy regime it is likely that virtually everyone emits subsistence emissions to some degree (Shue 2014, 197-98; Baatz 2014). The reason it may not look that way is that we equate subsistence emissions with an exemption from mitigation burdens, which as I shall explain shortly is a mistake. The concept of subsistence emissions as such, however, is uninterested in whether someone is rich or poor; what matters is simply whether emissions are necessary to satisfy a vital interest. Once we see this, we realize that most rich people are also subsistence emitters to a certain degree.

3. The Moral Force of Subsistence Emissions

Having defined subsistence emissions, let us now consider their moral force. What are the normative consequences of the fact that a set of emissions is of the subsistence variety? One answer connects to moral permissibility. Call this:

The Permission Claim. People are morally permitted to emit subsistence emissions.

Climate ethicists almost universally endorse the permission claim. The underlying thought is that people have a right to produce subsistence emissions because there is a limit to the costs we can expect people to absorb for the sake of others, even when we are considering acts are harmful and in principle avoidable (see e.g. Tadros 2011, 127-38). To expect people to refrain from subsistence emissions would be unreasonable because it would require enormous and perhaps fatal sacrifices on their part.⁹ We might admittedly wonder whether people have a moral permission to satisfy their vital interests when this comes at the price of frustrating other people's vital interests, which seems to be case for climate change. Since emissions jeopardize vital interests just as much as they protect them, perhaps we need to parse the interests more finely, such that people are not permitted to satisfy their decency interests if it comes at the price of frustrating other people's survival interests.¹⁰ But while parsing vital interests in this way would certainly have radical implications for the allocation of emissions rights, it would not question the right to satisfy survival interests, nor the more general point that subsistence emissions have priority over non-subsistence emissions.

The permission claim, then, is plausible. Some, however, ascribe a further normative power to subsistence emissions, which connects to climate justice and more specifically the fair allocation of costs or burdens of combatting climate change. The thought here is that those who emit subsistence emissions should not be held responsible, financially or otherwise, for this. Call this:

⁹ Subsistence emissions should not be mixed up with *unavoidable* emissions. Some non-subsistence emissions are unavoidable, but while this makes them morally permissible, it obviously does not make them subsistence emissions. Conversely, many subsistence emissions are (at least in a literal sense) avoidable, but the relevant point is not that people *cannot* avoid subsistence emissions but that they *should not* be expected to.

¹⁰ For the logic here, see Shue (2020, 127-128) and Vanderheiden (2008, 243).
The Exemption Claim. People are under no moral obligation to take mitigation burdens because of emitting subsistence emissions.

"Mitigation burdens" here refers to acts that contribute to halting or reducing the concentration of GHGs in the atmosphere, where performing such acts involves some cost or effort.

The exemption claim is seldom singled out for scrutiny, but climate ethicists often endorse it implicitly or explicitly. The claim is visible, for example, in Rao and Baer's idea that subsistence emissions constitute "a morally justified threshold for exemption from mitigation burdens" (Rao and Baer 2012, 659) as well as in Vanderheiden's remark that "assessments of liability" cannot be made against those who emit subsistence emissions since "all persons have valid claims to emit GHGs up to the survival threshold" (Vanderheiden 2008, 243).¹¹ The idea is that no one should be expected to partake in efforts to mitigate climate change because of having emitted subsistence emissions.

The exemption claim faces questions regarding the composition of a person's overall emissions. Suppose half of a person's emissions are subsistence emissions and the other half is not. I take it that the idea would then be that people could be potentially be asked to reduce the latter emissions but not the former. It would not seem plausible if a person could not be asked to mitigate at all just because *some* of her emissions are of the subsistence variety. If virtually everyone emits subsistence emissions to some degree, this would lead to the absurd conclusion that virtually no one could be asked to reduce emissions. But the more precise idea is that actors should be exempted for emitting *insofar as* they have emitted subsistence emissions. One way of living up to this constraint would be to allocate responsibilities to mitigate based strictly on one's share of total non-subsistence emissions (Vanderheiden 2008, 71). The implication would be that a person who only emits subsistence emissions would have no moral obligation to reduce emissions at all.

The exemption claim may look convincing at first glance, but it should be rejected. The problem is that it just does not follow from the fact that people are morally permitted to emit GHGs that they should be exempt from mitigation burdens. In this section, I show that if we should give priority to vital interests in the first place, then the only emitters that should be exempt from mitigation burdens are those whose vital interests would be frustrated if they were to take such burdens.

¹¹ Rao & Baer (2012) think that a threshold of exemption is provided by what they call "decent living emissions", but this does not change the underlying point that vital emissions should be exempted from mitigation burdens.

3.1 Qualified Moral Permissibility

Let us begin by considering the permission claim, because this is no doubt what is thought to underpin the exemption claim. The argument for the permission claim can be stated as follows:

- (1) People are morally permitted to satisfy their vital interests.
- (2) Subsistence emissions are needed to satisfy vital interests.
- (3) Therefore, people are morally permitted to produce subsistence emissions.

We might think that if this argument is sound, then the exemption claim follows, but this is a flawed inference. We cannot automatically move from the idea that people are permitted to produce emissions to the conclusion that they should be exempt for doing so. This is because there are two different senses of being morally permitted one could invoke. Both agree that a moral permission to φ means that no moral duty is violated in φ -ing, but they differ as to whether doing something permissible may generate subsequent compensation duties or not. On what we may call the *unqualified permissibility view*, if we are morally permitted to φ , then we have no duty of compensation for φ -ing afterwards. On what we may call the *qualified permissibility view*, by contrast, we may have duties to compensate others for having φ -ed even though we were morally permitted to φ .¹² The distinction between these two views means that the argument for the permission claim is underspecified. Depending on which sense of "permissibility" we invoke, we could spell out its key premise in two different ways:

- (1*) People are morally permitted to satisfy their vital interests, and they are under no compensatory duties for doing so afterwards.
- (1**) People are morally permitted to satisfy their vital interests, but they may be under compensatory duties for doing so afterwards.

The exemption claim only follows automatically if we understand the argument as referring to unqualified moral permission (1*).

¹² The term "qualified" is not ideal since the permission *to* φ is still fully there. But it is better than "conditional", because that would suggest that the permission itself depends on compensation later on. That would be too strong since it would mean that people are always forbidden from φ -ing when they can foresee that they will be unable to compensate for doing so afterwards. The point I want to make is that people can incur compensatory duties despite acting permissibly; and if they fail to discharge these duties despite being able to, then they commit a *separate* wrong.

I argue that the qualified permissibility view is sometimes correct when it comes to emitting subsistence emissions. More specifically, qualified permissibility holds for any actor who is able to compensate others for producing subsistence emissions without frustrating their vital interests. This means that it is a mistake to equate subsistence emissions with an exemption from mitigation burdens. We should instead say that while everyone is permitted to produce subsistence emissions, some are nevertheless under a duty to compensate others for this.

What speaks in favour of the idea that people can be permitted to do something in a qualified way? Ultimately, the compelling moral intuition that actors may be permitted to do one thing even though they *also* have a moral duty to set things right afterwards. This is nicely brought out by Joel Feinberg's famous "cabin case":¹³

Suppose that you are on a back-packing trip in the high mountain country when an unanticipated blizzard strikes the area with such ferocity that your life is imperiled. Fortunately, you stumble onto an unoccupied cabin, locked and boarded up for the winter, clearly somebody else's private property. You smash in a window, enter, and huddle in a corner for three days until the storm abates. During this period you help yourself to your unknown benefactor's food supply and burn his wooden furniture in the fireplace to keep warm (Feinberg 1978, 102).

Feinberg thought that given the danger you would be "justified in doing all these things" even though it would infringe "the clear rights of another person" (ibid.). But he also thought that this does not mean that you could just walk away from the situation without taking steps to redress the cabin owner. Feinberg wrote that,

We would not think it inappropriate to express our gratitude to the homeowner, after the fact, and our regrets for the damage we have inflicted on his property. More importantly, almost everyone would agree that you owe *compensation* to the homeowner for the depletion of his larder, the breaking of his window, and the destruction of his furniture (ibid. 233).

This just seems plausible. People no doubt have a moral right to break into cabins to survive blizzards, but it would be odd to think that there is no moral requirement to undo the losses they inflict in the process. There is clearly something to regret about how the backpacker had to act to get out of harm's way. If the backpacker were to

¹³ Feinberg used the cabin case to illustrate that there can be justified rights-infringements. My argument can, but need not, be read as saying that subsistence emissions can be a justified infringement of others rights not to have their vital interests jeopardized by climate change.

think that he owed no compensation just because he was permitted to save his own life, then he would be guilty of conflating two issues: whether it is permissible to act in a certain way and whether it is permissible to walk away without undoing losses brought about in the process.

The cabin case strongly suggests, then, that moral permissions can be qualified.¹⁴ Are they qualified in the case of emitting subsistence emissions? It seems so. Since emitting GHGs contributes to harm and depletes the atmosphere's ability safely to absorb further emissions, it is plausible that there can be cases where a person owes compensation for emitting despite the fact that it was necessary to satisfy vital interests. Indeed, according to the sufficientarian morality that underpins the basic right to produce subsistence emissions, we can say something stronger and more precise, for the implication of this view is that "qualified" moral permissibility holds for all situations where emitters could compensate for their subsistence emissions without jeopardizing their vital interests. The notion of an "unqualified" permission is strictly reserved for those who would be unable to satisfy their vital interests if they were to compensate for emitting.

To drive home these points, consider the position of someone like me, a well-paid citizen of the developed world. Some of my emissions are no doubt subsistence emissions, and it is plausible that I should not be expected to refrain from them. I have a right, for example, to adequate housing and nutrition and thus a right to produce whatever GHG emissions are necessary to achieve those things.¹⁵ Yet that does not mean that I have an *unqualified* right to produce subsistence emissions. Since the emissions threaten other people's vital interests, surely I ought to compensate for them considering that I have plenty of economic "spare capacity" before I reach a point where asking me to compensate would jeopardize a decent minimum, let alone my survival.

The concept of subsistence emissions, then, has variable moral force. It only exempts when we are dealing with emitters who would fall below, or fail to reach, the decent moral minimum if they were to compensate for their emissions. The reason this point has been overlooked is probably that we tend to think that only poor people produce subsistence emissions. But what does a duty to "compensate" for one's subsistence emissions involve more exactly? The answer here is no different from how we normally think about compensatory duties because of emitting GHGs. The idea is that we can compensate for our emissions by *offsetting* them, that is, by preventing or reducing emissions by an equivalent amount else-

¹⁴ To be explicit, I am not suggesting that they are *always* qualified.

¹⁵ I use "produce" in a loose sense here. I seldom produce the emissions directly but rather consume goods and serviced which *embody* emissions. The emissions from transporting my food is an example of this.

where. This may involve enhancing natural or artificial carbon sinks, or lowering the emissions of someone else. For example, we can plant trees, invest in carbon capture and storage, or help pay for the premature closure of a coal plant. Such actions can be compensatory because they can ensure that our net contribution to the atmospheric stock of GHGs is zero. The moral point is that people who are able to do this without jeopardizing their vital interests ought to do so, whether or not the emissions were of the subsistence variety.

It is worth underlining three further things about the argument. First, the way it uses the term "compensation" does not align with much of the climate justice literature. It is common in the literature to reserve "compensation" for acts which redress victims of materialized climate harm; what in climate policy circles is known as "loss and damage" (Caney 2012; Heyward 2013; Page and Heyward 2017). But while loss and damage represents a distinct climate policy, we can also say that actors engage in "compensation" when emissions have not contributed to materialized harm (Duus-Otterström and Jagers 2012). For example, when an actor fully offsets his or her emissions, it is perfectly appropriate to say that the actor has "compensated" for the fact that he or she produced some emissions that ended up in the atmospheric stock of GHGs. It is in this looser sense I use "compensation".

Second, the compensatory duty is *pro tanto* and not categorical. There may be other moral considerations which make it the case that, all things considered, people do not have a moral duty to compensate for their emissions or indeed that they *should not* do so. This point is a simple consequence of recognizing that climate change and climate justice is not all that matters from a moral perspective.

Third, in order to compensate for one's emissions in the sense used above, it is important that the compensation is prompt. It will not do if people compensate for their emissions years later, because then the emissions will have been forcing the climate in the intervening period. Spelling out this timing requirement is difficult, but the general message is that compensation should occur as quickly as possible. Indeed, the best would probably be if people offset their subsistence emissions beforehand, so that their contribution to the atmospheric stock of GHGs would be negative in the period between the emissions and the compensation.

3.2 The Implications of Qualified Permissibility for Climate Justice

Having defended the argument, let us now see how it affects the burden sharing discussion. Subsistence emissions have always had a radical potential since they rest on a logic which is fundamentally incompatible with the idea of keeping emissions within a safe carbon budget. The risk is that when each person's subsistence emis-

sions are added together, the result is significant climate change, at least now that much of Earth's absorptive capacity has already been used up. If so, the upshot of the right to produce subsistence emissions is a moral right to cause harmful climate change.

Now it seems true that the right to produce subsistence emissions does give rise to this conflict, and absent the kind of technological and economic developments that would allow humanity to sever the link between GHG emissions and the satisfaction of vital interests, perhaps we must begin the work of qualifying the right to produce subsistence emissions itself, in light of the other and perhaps weightier vital interests that climate change threatens. This would involve ranking the moral weight of different vital interests as well as, dauntingly, sorting out the importance of numbers. For example, does it matter if those who produce subsistence emissions threaten the vital interest of a much larger number of people? These are difficult and contentious issues, but the idea of qualified permissibility at least lessens the size of the challenge. It reminds us that even though people have a right to produce subsistence emissions, many will also have a duty to contribute to the effective management of the climate problem as compensation for this.

Whether the right to produce subsistence emissions adds up to significant climate change depends, of course, on the number of people who merely possess a qualified permission to emit. I shall not speculate about this question. I will just note that it is plausible that anyone who is at least reasonably rich—such as an average citizen of high-income or higher middle-income countries—has the spare capacity to compensate for subsistence emissions without frustrating their vital interests. Unqualified permissions to emit is arguably reserved for the poor, whose vital interests generally *would* be frustrated further if they were required to offset their emissions. In the end, then, we end up in a familiar place, where what is at stake is the right to develop among the world's poor. Yet we now see that what explains their being exempt is not that they must emit to satisfy their vital interests, but that partaking in burden sharing would be incompatible with their reaching, or maintaining, a decent human minimum. In this way, the moral force of subsistence emissions adheres to a sufficientarian version of the Ability to Pay Principle (Roser & Seidel 2017, 140–149).

How could one approximate the cut between qualified and unqualified permissibility in practice? The blueprint is arguably laid out by accounts such as the Climate Equity Reference Framework (Holz, Kartha, and Athanasiou 2018) and the Greenhouse Development Rights Framework (Baer 2010), which seek to calculate fair climate mitigation targets for different countries. These accounts exclude individuals whose income is under a certain level from the calculation of national mitigation burdens. According to the Greenhouse Development Rights Framework, for example, individuals whose annual purchasing power parity adjusted income is below \$7,500 should be exempt from climate mitigation burdens. We may debate whether this particular number is the correct threshold, but the basic idea of offering exemptions based on income could track the distinction qualified and unqualified permissions relatively well.¹⁶

4. Objections

Let us now consider some objections, beginning with two kinds of scepticism that threaten to undercut the argument completely. First, there is scepticism about whether people act wrongly in emitting GHGs. Some think that the individual's emissions are so tiny compared to the total that they make no morally relevant difference.¹⁷ They would reject that people do anything wrong in emitting GHGs, at least because it inflicts harm on others. Second, there is scepticism about whether it is practically possible to offset one's emissions in a reliable way. There are well-documented problems with the voluntary offsetting market that is presently in place, and the risk is that these problems will plague any attempt to compensate for one's emissions by preventing someone else from producing the same amount.¹⁸ If correct, either form of scepticism would stop the argument from getting off the ground: the former by rejecting that emitting can be grounds for compensation; the latter by rejecting that people *could* compensate for emitting even if they ought to do so in principle.

Both sceptical perspectives can safely be put to one side here. The debate about individual climate obligations, while interesting, is orthogonal to the debate about subsistence emissions, because if individual emissions are so small to be morally inconsequential, then this would question individuals' duty to reduce *any* emissions and not just subsistence emissions. For the purposes of discussing the role of specifically subsistence emissions in climate justice, we are entitled to assume that individuals can act wrongfully in emitting GHGs, whether this is because they contribute to harm, appropriate an unfair share of the overall emissions budget, or

¹⁶ The Climate Equity Reference Framework and the Greenhouse Development Rights Framework also exempt the fraction of a country's emissions that is due to consumption under the threshold. This is evidence that their key concern is with ability to pay. It is unclear, however, whether they reject the permission claim or the idea that emissions over the development threshold can be subsistence emissions.

¹⁷ Hiller (2011) calls this the claim of "individual causal inefficacy". This claim was famously defended by Sinnott-Armstrong (2010). For an excellent overview of the debate, see Fragnière (2016).

¹⁸ For an overview of the ethical issues of carbon offsetting, see Hyams and Fawcett (2013). The thorniest theoretical issue is arguably additionality: how can we be sure that our offsetting truly prevents emissions as opposed to merely paying for "preventing" emissions that would not have occurred anyway?

something else. Scepticism about the feasibility of compensation, meanwhile, can be put to one side because it points to contingent features of offsetting which may improve in the future, and because it overlooks the possibility of investing in negative emissions technology, which seems less affected by the sort of moral hazard and baseline problems that plague traditional offsetting.

Having duly noted the sceptical broadsides, let us now consider objections that are more internal to the argument. A first objection is that it is not clear what it means to "compensate" for emissions in a world in which everyone's emissions eventually must come down to zero. Suppose, for example, that the only emissions that occur at some point are subsistence emissions. How could an emitter then compensate for emitting, as per the idea of qualified permissibility? It seems that compensation would here only be possible if other people—the target of the offsetting—were to forgo satisfying *their* vital interests.

This objection is not really an objection to the argument as much as a challenge to its scope of application. It might well be that unqualified moral permissions would be the norm in a world where there is only subsistence emissions. If people could not discharge their duty of compensation (because this would require that others forgo protecting their own vital interests) then they would not bound by this duty to begin with. But while the objection is correct to question the practical relevance of qualified permissibility, it overstates the problem of compensation because it assumes a particular understanding of what it means to compensate for one's emissions, namely, reducing *someone else's* emissions. Yet as mentioned offsetting can also take the form of enhancing natural or artificial carbon sinks, such as when we help pay for reforestation or the deployment negative emissions technologies, and these things do not require that other people reduce their emissions. It is not true, then, that the duty of compensation could not be discharged in a world where there are only subsistence emitters.¹⁹

A second objection is that the account I have offered is too strict because it assigns a duty of compensation to everyone that can compensate without jeopardizing his or her vital interests. We often think that people are entitled to emit their "fair share" of GHGs without accruing any compensatory duties, where a fair share includes more than just subsistence emissions. So we might feel that the bar for unqualified moral permission is set too high if the only people who would be permitted to emit in an unqualified sense would be those who are, or would become, intolerably badly off if they were to shoulder climate burdens.

This is a fair objection that necessitates yet another distinction, because my

¹⁹ This conclusion would of course be clearer still if people could compensate for having emitted in ways other than mitigation. I set other kinds of climate burdens aside here, but for the case for fungibility across climate burdens, see Duus-Otterström & Jagers (2012).

argument could be read as expressing a strong or a weak thesis. The *strong* thesis holds that the only people who have an unqualified moral permission to produce subsistence emissions are indeed those who are, or would become, intolerably badly off if they were to shoulder climate burdens. It does not matter if someone has emitted very little in the past; anyone who has the capacity to compensate for emitting ought to do so, even if the emissions in question are necessary adequately to satisfy vital interests. The *weak* thesis, by contrast, says that other actors may also have an unqualified moral permission to emit, for example because they are still emitting within their fair share. Neither thesis questions that the destitute are able to emit subsistence emissions without accruing duties of compensation; the difference lies in whether the destitute are *the only ones* able to do this.

Where we come down on the choice between the strong and the weak thesis depends on whether we see climate change as a problem where "harm avoidance justice" takes precedence over "burden sharing justice" (Caney 2014). I have no interest in settling this debate here. It is enough to point out that there are coherent arguments for both positions. What speaks in favour of the strong thesis is the urgent need to minimize further climate change. Given how serious the climate problem has become, it might indeed be the case that exemptions from responsibility must be highly conservative.²⁰ What speaks in favour of the weak thesis, meanwhile, is our sense of fairness. We might feel that unqualified moral permissions should extend all people who so far have emitted little, especially since others were able to produce even luxury emissions without being expected to compensate for this in the past.

A third objection is that the distinction between subsistence and non-subsistence emissions is irrelevant because *everyone* is morally permitted to produce *any* emissions in a "qualified" way. The thought here is that even the most frivolous luxury emissions must be permissible if the emitter fully compensates for them. Thus, there is nothing special about producing subsistence emissions. People should compensate for any emissions if they are able to do so without frustrating their vital interests, but as long as they do compensate, it does not matter if the emissions protect the means of subsistence or sheer luxuries.

The response to this objection can take several forms. One option is to just accept that all emissions are morally permissible as long as the emitter compensates for them. Indeed, if offsetting is fully successful, it is an open question whether the emitter has "emitted" anything in the relevant sense since the net contribution to the stock of GHGs is zero. Yet we might also respond in ways that preserve the

²⁰ For the view that all reasonably avoidable emissions are morally wrong, see Hiller (2011) and Broome (2012).

special significance of subsistence emissions. First, we might permit weaker forms of compensation than full offsetting. If so, one possibility is to hold that these weaker forms apply to subsistence emissions, whereas those who produce nonsubsistence emissions have a duty to compensate in the stronger sense (i.e., ensure that the net contribution to the atmospheric stock of GHGs is zero). Secondly and more importantly, there are good reasons to question whether all emissions are permissible as long as there is compensation. This is partly because of epistemic problems-it is difficult to know which acts truly are offsetting, and we should not be cavalier about our ability reliably to compensate for them-but mainly because there is a morally better outcome, namely one where we refrain from emitting but engage in indirect mitigation anyway. That is, rather than saying that people are permitted to produce emissions as long as they compensate for them, we might want to say that people should avoid producing emissions and engage in emissions reductions elsewhere. This approach, however, would not be plausible for subsistence emissions precisely because one must produce such emissions to satisfy vital interest. So the idea of qualified permissibility does seem to apply to subsistence emissions in a distinctive way.

5. Conclusion

Climate change is fundamentally an energy problem, and the fundamental task facing humanity is to sever the link between energy and fossil fuels, enabling people to enjoy a good standard of living without jeopardizing the climate's stability. Yet until this link is severed, GHG emissions remain a by-product of producing energy, and the demand for such emissions outstrip their supply. This raises the question of how rights to emit should be allocated between people. The idea of subsistence emissions looms large in the way climate ethicists have answered that question. The thought is that people cannot reasonably be expected to give up emissions that are necessary adequately to satisfy their vital interests, and that this fact must be reflected in the way the remaining carbon budget is shared.

I have argued that the role of subsistence emissions in climate justice is more complicated than standardly thought. While there is little doubt that people have a right to *produce* subsistence emissions, this does not rule out that people can be under a moral duty to *compensate* others for this afterwards. Far from issuing a general exemption from responsibility, the basic morality that grounds our rights to satisfy our vital interests suggests that anyone who is able to compensate for emitting without frustrating their vital interests have a *pro tanto* moral duty to do so. Once we take this point on board, we see that the concept of subsistence emissions has less radical implications for climate change than some think. While people have

a right to produce whatever GHGs are necessary adequately to satisfy their vital interests, the aggregated result need not be runaway climate change considering that many people will have the spare capacity to compensate for those emissions.

References

Agarwal, Anitra, and Sunitra Narain. 1991. "Global Warming in an Unequal World". New Delhi: Centre for Science and Environment.

Baatz, Christian. 2014. "Climate Change and Individual Duties to Reduce GHG Emissions." *Ethics, Policy & Environment* 17 (1): 1–19.

Baer, Paul. 2010. "Greenhouse Development Rights". In *Climate Ethics: Essential Readings*, edited by Stephen M. Gardiner, Simon Caney, Dale Jamieson, and Henry Shue, 215-230. Oxford: Oxford University Press.

Brock, Gillian, and David Miller. 2019. "Needs in Moral and Political Philosophy". In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2019. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2019/entries/needs/.

Broome, John. 2012. Climate Matters. London: W.W. Norton.

Caney, Simon. 2009. "Justice and the Distribution of Greenhouse Gas Emissions1". *Journal of Global Ethics* 5 (2): 125–46.

----. 2012. "Just Emissions". Philosophy & Public Affairs 40 (4): 255-300.

----. 2014. "Two Kinds of Climate Justice: Avoiding Harm and Sharing Burdens". *Journal of Political Philosophy* 22 (2): 125–49.

Crocker, David. 2008. *The Ethics of Global Development*. Cambridge: Cambridge University Press.

Duus-Otterström, Göran, and Sverker C. Jagers. 2012. "Identifying Burdens of Coping with Climate Change: A Typology of the Duties of Climate Justice". *Global Environmental Change*, 22 (3): 746–53.

Feinberg, Joel. 1978. "Voluntary Euthanasia and the Inalienable Right to Life". *Philosophy & Public Affairs* 7 (2): 93–123.

Fragnière, Augustin. 2016. "Climate Change and Individual Duties". *WIREs Climate Change* 7 (6): 798–814.

Gough, Ian. 2015. "Climate Change and Sustainable Welfare: The Centrality of Human Needs". *Cambridge Journal of Economics* 39 (5): 1191–1214.

Griffin, James. 1986. *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Clarendon Press.

Heyward, Clare. 2013. "Situating and Abandoning Geoengineering: A Typology of Five Responses to Dangerous Climate Change". *PS: Political Science & Politics* 46 (01): 23–27.

Hiller, Avram. 2011. "Climate Change and Individual Responsibility". *The Monist* 94 (3): 349–68.

Holz, Christian, Sivan Kartha, and Tom Athanasiou. 2018. "Fairly Sharing 1.5: National Fair Shares of a 1.5 °C-Compliant Global Mitigation Effort". *International Environmental Agreements: Politics, Law and Economics* 18 (1): 117–34.

Hyams, Keith, and Tina Fawcett. 2013. "The Ethics of Carbon Offsetting". *WIREs Climate Change* 4 (2): 91–98.

Mancilla, Alejandra. 2019. "The Human Right to Subsistence". *Philosophy Compass* 14 (9): e12618.

Miller, David. 2007. *National Responsibility and Global Justice*. Oxford: Oxford University Press.

Page, Edward A, and Clare Heyward. 2017. "Compensating for Climate Change Loss and Damage". *Political Studies* 65 (2): 356–72.

Rao, Narasimha D., and Paul Baer. 2012. "Decent Living' Emissions: A Conceptual Framework". *Sustainability* 4 (4): 656–81.

Roser, Dominic, and Christian Seidel. 2017. *Climate Justice: An Introduction*. London: Routledge.

Shue, Henry. 1993. "Subsistence Emissions and Luxury Emissions". *Law & Policy* 15: 39–60.

----. 2014. *Climate Justice. Vulnerability and Protection.* Oxford: Oxford University Press.

———. 2019. "Subsistence Protection and Mitigation Ambition: Necessities, Economic and Climatic". *The British Journal of Politics and International Relations* 21 (2): 251–62.

----. 2020. *Basic Rights: Subsistence, Affluence, and U.S. Foreign Policy*. Princeton: Princeton University Press.

Sinnott-Armstrong, Walter. 2010. "It's Not My Fault. Global Warming and Individual Climate Obligation". In *Climate Ethics: Essential Readings*, edited by Stephen M. Gardiner, Simon Caney, Dale Jamieson, and Henry Shue, 332–46. Oxford: Oxford University Press.

Tadros, Victor. 2011. *The Ends of Harm: The Moral Foundations of Criminal Law*. OUP Oxford.

Vanderheiden, Steve. 2008. *Atmospheric Justice: A Political Theory of Climate Change*. Oxford University Press.

Julia Mosquera¹

Climate Change, Corrective Justice, and Non-Human Animals²

While duties of corrective justice in the case of harms to human beings are widely accepted, there is almost non-existing discussion on whether these duties should be extended to non-human animals, even though animals are also wronged by our actions, and sometimes even to a greater extent. Very few maintain that we owe nothing to non-human animals, especially to those in need. Many agree that we have at least some duties of justice to non-human animals, for example, of distributive justice (based on priority (see Regan, 1984; Holtug, 2007), equality (see Horta, 2016; Persson and Vallentyne, 2005), or sufficiency (Crisp, 2003). In this paper I explore how corrective justice can be extended to non-human animals. The paper investigates why certain normative attitudes are not necessary in order to count as a claimant of corrective justice. I suggest which responses can count as fitting when it comes to discharging duties of corrective justice to non-human animals. I also discuss who should count as a candidate for having claims of corrective justice, that is, to whom we owe duties of corrective justice (i.e. ecosystems, species, individual non-human animals, or rather, groups of non-human animals). Finally, I raise concern about the

¹Institute for Futures Studies; julia.mosquera@iffs.se.

² I am grateful to the participants at the Climate Change and Corrective Justice Workshop at IFFS (June 2019) and at the Animals and Climate Change Workshop at IFFS (December 2020) for helpful comments to this paper. Financial support from Riksbankens Jubileumsfond (Sweden) (Grant No. M170372:1) is gratefully acknowledged.

difficulty of determining a baseline for the correction of wrongs to nonhuman animals that live in the wild, given that their original situation is far from idyllic. This has important implications in deciding how to correct wrongs like those of climate change.

1. Introduction

Climate change harms both human and non-human animals. It is uncontroversial that humans harmed by climate change are owed duties of corrective justice. It is commonly accepted that non-human animals are owed at least some duties of justice, too. Many would agree that we have duties of beneficence, as well as duties of distributive justice to non-human animals, whether the latter should be understood in terms of priority, equality, or sufficiency. If so, there seems to be a case for duties of corrective justice to apply to non-human animals, too, especially in the context of climate change. This is a question of high relevance, both theoretical and practical.

Climate change harms both human and non-human populations. The IPCC recognises these impacts: "On land, impacts on biodiversity and ecosystems, including species loss and extinction, are projected to be lower at 1.5°C of global warming compared to 2°C" (IPCC, 2018).

Recent scientific evidence points to climate change as a direct driver of the current loss of biodiversity and further risks of it, including the animal kingdom. The latest IPBES report presented the following conclusions regarding marine and terrestrial biodiversity:

... many studies estimate the fraction of species at climate change related risk of extinction is 5% at 2°C warming, rising to 16% at 4.3°C warming {xx}. Climate change and business-as-usual fishing scenarios are expected to worsen the status of marine biodiversity (*well established*) {4.2.2.2, 4.2.2.3.1}. Climate change alone is projected to decrease ocean net primary production by between 3 and 10 per cent and fish biomass by between 3 and 25 per cent (in low and high warming scenarios, respectively) by the end of the century (*established but incomplete*) {4.2.2.2.1}. (...) The average abundance of native species in most major land-based habitats has fallen by at least 20%, mostly since 1900; The numbers of invasive alien species per country have risen by about 70% since 1970, across the 21 countries with detailed records; and the distributions of almost half (47%) of land-based flightless mammals, for example, and almost a quarter of threatened birds, may already have been negatively affected by climate change. (IPEBES, 2019: 7)

It could be further argued that climate change will harm non-humans more than humans given their greater number and already low quality of life. With respect to lower-welfare, it has been recently suggested that one should reject the 'idyllic view of nature'. According to this argument, if we take animals that live in the wild, suffering prevails over pleasure. Among the reasons for this harmful situation are the natural catastrophes these animals are exposed to, the early death of millions of new-born animals born from species with large progeny of which only a reduced number of descendants survive, and the dynamics of predation to which many of them are exposed (Horta, 2013: 113–125).

With respect to numbers, animals living in the wild exceed in many orders of magnitude the present number of human beings and nonhuman animals under exploitation (Tomasik, 2009). And free-living animals are more vulnerable to climate change due to their high exposure to the environment in which they live and their low ability to cope with it throughout their lifetimes. While some species as a whole may be able to adapt to climate change, individual animals commonly cannot.

Climate change is mostly humanly caused and therefore gives humans some duties. These duties have been normally categorised as duties of mitigation, adaptation, and correction or compensation for climate change. This paper focuses on the latter kind of duties. Corrective justice focuses on identifying the victims and the perpetrators of a wrong, and in determining what

Very few would deny that we have duties of corrective justice towards human beings who have been wronged, such as in the context of climate change. The search for principles of justice that aim at distributing the costs of reducing mitigation and the costs compensation, such as some version of the Polluters Pay Principle (PPP) or Rawls' Just Savings Principle, is evidence of the acceptance of the role of corrective justice in the context of climate change.

While duties of corrective justice in the case of the harms of climate change to human beings are widely accepted, there is almost non-existing discussion on whether these duties should be extended to non-human animals, too, given they are also directly wronged by our emissions, and possibly to a greater extent.

Very few would maintain that we owe nothing to non-human animals, especially to those in need. The ground for duties of justice to non-human animals has already been established. Many agree that we have at least some duties of justice to nonhuman animals, for example duties of distributive justice. According to cosmopolitanism, justice is owed to all beings in the world who have the requisite psychological make-up and existential status (e.g., Pogge, 1992). Some cosmopolitans take the required make-up to be rational agency, which would exclude non-human animals and children. To avoid the conclusion that we do not have duties of justice to neither animals nor children, others adhere to cosmopolitan views for which sentience or having the relevant interests is the requisite make-up. According to these views, we have some duties of distributive justice to both children and non-human animals.

In the case of duties of distributive justice, some argued that we have duties of priority to non-human animals, based on they idea that they some of these are worse

than we are and that therefore helping them constitutes a greater benefit (see Regan, 1984; Holtug, 2007). Others have suggested instead that our duties to nonhuman animals are duties of equality (see Persson, 1993; Vallentyne, 2005; Horta, 2016). And some others have instead proposed that we should provide to nonanimals according to some sufficientarian principle (Crisp, 2003).

In this paper I explore the extent to which corrective justice can be extended to non-human animals and, if so, in which way. The paper aims at exploring some of the philosophical questions that this issue raises, such as to which extent we owe compensation to individuals who, while having been wronged, are not able to recognise acts of corrective justice as such, and not simply as a mere positive reward; as well as the moral and metaphysical status of compensation in such cases. The conclusions reached here can shed light on the context of duties to other individuals who might not share our understanding and practices of these duties.

Finally, and given the current state of affairs and prospects, some have raised the doubt about whether climate change should be understood as a pessimistic or as an optimistic scenario for non-human animals (Faria & Paez, 2020; Palmer, 2011). Under the pessimistic scenario, the average wellbeing of free-living individuals will be lower due to climate change. The pessimistic scenario assumes that the effects of climate change will be overall bad for human beings. This seems like the most obvious scenario. Under the optimistic scenario, the average wellbeing of free-living individuals will be higher due to climate change. Faria and Paez motivate the optimistic scenario in the following way:

This scenario [the optimistic scenario] is not altogether implausible once we reject the idyllic view of nature. The effect of climate change in certain ecosystems may lead to a reduction in the quantity of free-living animals that they are able to support. As explained, we have compelling reasons to believe that the life of most of these animals is net negative, containing more suffering than positive experiences. If climate change causes a number of animals not to exist who would have otherwise been born, that reduces the amount of suffering in the wild. If this was the main effect of climate change in ecosystems, the overall result would be a net reduction in animal suffering. In terms of their own well-being, then, climate change would be good for free-living animals. (Faria & Paez, 2020)

Given the lack of knowledge on the effects of climate change for non-human animals, some have argued that that we should be uncertain about whether we should set for the pessimistic or the optimistic scenario. Claire Palmer (2017) describes some of the reasons to be uncertain about this scenario: This future world would be packed with trillions of living things, though if we look far enough into the future, barely any of them would be the same living individuals that currently exist. Suppose we now look through this individual-focused lens at a future world with climate change. This world is still filled with trillions of living things, distinct from those currently alive. But many, or most, of the individuals in this future world with climate change are different individuals from those that would have existed in the alternative future world without climate change. Where the same species exists in both future worlds, particular genetic individuals almost certainly differ (as a changing climate, for instance, affects which individuals mate and produce offspring). And in the climate-changed world, there are likely to be fewer-or perhaps no -individuals of some species. But individuals of other species are predicted to be more numerous, flourishing, and found in new locations; evidence of this process already exists (...) Exactly what these impacts will be is empirically uncertain, but they may include changing the number of individuals, the existence of different individuals, harm or death to some individuals, and the bringing into being or benefitting of other individuals. (Palmer, 2017: 106-7).

The uncertainty argument about what kind of scenario climate change represents for non-human beings is a significant issue for wellbeing assessment of the consequences of climate change. While the species mix might change, it is unclear whether the number of total individuals will be smaller. Also, while some species will disappear, others will take over. Because of this, what matters would be rather whether fewer individuals with sophisticated psychological characteristics disappeared, without them being substituted by other species of the same or comparable psychological complexity or complexity to feel pain or pleasure.

The argument developed in this paper is compatible with this uncertainty. The uncertainty argument does not undermine the goal of this paper. While it might be true that we should ascribe some credence to the scenario in which climate change results in an overall positive outcome for non-human animals due to their already very low level of welfare, this is compatible with determining the duties that we have to those animals already harmed by climate change, and that will continue to be harmed. Corrective justice applies to harmed individuals, or to individuals at risk of being harmed (McKinnon, 2009) and therefore is indifferent to whether an event like climate change may eventually lead to a greater net reduction in animal suffering due to the prevention in the existence of animals that would have otherwise had wrongful lives.

2. Corrective justice and the argument from normative attitudes

Corrective justice is the branch of justice that focuses on ensuring that any party liable for harm done to the rights and/or interests of another make reparations (often in the form of compensation) to the victims so as to redress the imbalance of justice between them (cf Mckinnon, 2008).

Corrective justice must be distinguished from other branches of justice, such as distributive justice. Distributive justice consists in determining and implementing a just allocation of goods between a set of individuals. Corrective justice connects a party with another in virtue of an action performed by the liable party, and these connections might not be true of the parties' relations with other members of their society. The idea is that the perpetration of an injustice destroys the presumed fairness that characterises ideal interactions between individuals, and that after an injustice is perpetrated, something needs to be done in order to make up for the moral harm caused and to restore the original justice setting.

Jules Coleman (1995) has identified some conditions that must be met in order for an agent to have a claim of corrective justice against another party. According to her, a person A suffering harm has a claim against a (putatively liable) person B under corrective justice if and only if

- (1) the harm to A is a result of human agency (as opposed to, for example, the forces of nature);
- (2) the harm to A is connected in a normatively significant way to B's actions (perhaps B caused the harm through her agency, or is otherwise responsible for it).
- (3) The appropriate form of A's claim against B takes the form of a claim for repair or rectification (often understood in practice as a claim for compensation).

Non-human animals harmed by climate change seem to meet the requirements specified by Coleman. (1) The climate change harms that non-human animals experience is a result, in great part, of human agency. (2) The climate change harms that non-human animals experience is connected to a certain party's actions, namely polluters, either by direct agency or some other source of responsibility. (3) The appropriate form of the non-human animal's claim against polluters takes *in*

some circumstances the form of a claim for repair or rectification. I will say more about this last point in the next section.

It must be noticed that Coleman's account does not specify any further requirement that for an agent to have a claim of corrective justice against another agent that agent must have certain normative attitudes, such as being able to understand the nature of wrongs, and well as the nature of compensations. And while these attitudes are present in the case of most human beings, they are not in the case of most non-human animals. Call this the *argument from normative attitudes*.

Although advocates of corrective justice for non-human animals themselves, Claire Palmer and Elizabeth Cripps raise worries similar to the one expressed by the argument from normative attitudes. Palmer alludes to the fact that animals would not be aware of the fact that they are being compensated when discussing the potential gains of compensating nonhuman animals for past harms committed to them:

(...) there are some standard gains from reparations that coyotes cannot experience. They cannot gain psychologically from knowing that reparation is reparation; anything like an *apology* or a memorialization would be wasted on them. They lack concepts of justice, bear no grudges against either perpetrators or beneficiaries, and seek no satisfaction from either. (...) They cannot be resentful, blame others, campaign for change, nor seek retribution. And there is no reason (unlike in the human case) to prefer to seek reparation from perpetrators on account of the satisfaction that this would give the victims, nor to be concerned about the effects on social disharmony if reparation is not extracted from the perpetrators. (Palmer, 2010: 104)

Cripps points to a similar worry:

(...) it is not clear that this [compensatory responses to nonhuman animals] would play the same role as compensation as in the human case. Among humans, there is an element of shared recognition of a wrong—at least arguably, a significance to being offered compensation over and above its practical impact—which does not apply here. (Cripps, 2013: 107-8).

I believe the argument from normative attitudes, in the way expressed by Palmer or Cripps, is not a decisive argument against a potential account of corrective justice for nonhuman animals. Justice gives rise to demands that go beyond individual's awareness and understanding of these, and injustice gives rise to claims, even if those who are unjustly treated lack the normative attitudes necessary to recognise a wrong a wrong as such, or the correction of this. Let me elaborate on this.

First, it is far from obvious that *no* nonhuman animal under any circumstance can recognise wrongs. This claim contrasts with recent research on behavioral and cognitive research that seems to show that nonhuman animals are expected to be treated fairly. According to Bekoff & Pierce (2009), certain species of nonhuman animals possess some sort of proto morality. They develop something similar to our sense of justice through prolonged relations with human beings. The same intuition is shared by Fran the Waal & Brosnan (2003) and other evolutionary biologists who argue that certain monkeys reject that they are treated unequally to others.

That nonhuman animals expect to be treated fairly would seem to show that animals do not only suffer when they are inflicted with pain, but they also suffer when their needs and desires are unjustly not met. Furthermore, a considerable number of nonhuman animals will be able to experience a rebuild of trust from a compensatory arrangement performed by their perpetrator.

This is especially true for farm and companion animals who live their everyday lives with humans and are dependent upon them. For example, it is also true for nonhuman animals with higher cognitive capacities such as gorillas. Animals with higher cognitive capacities are able to experience a rebuilding of trust after a wrong is committed to them if the wrongdoer modifies her attitude in the right way so that the animal that was wronged in the past can abandon the expectation that a similar wrong will be committed again in the future by the old offender—or by any other human similar to the offender, or who fulfills a similar role as the offender used to have with this animal.

Second, it is at least not obvious whether these normative attitudes are necessary for a response to wrongdoing to constitute a correction to an injustice. Let us assume, for the sake of the argument, that nonhuman animals lack the aforementioned normative attitudes. The concerns raised by Palmer and Cripps encapsulate two different ideas: the idea that the value of corrective justice relies, at least partly, on (1) the shared recognition of a wrong, and (2) the recognition of the significance of being offered compensation. According to these, if a wronged party—in our case, a nonhuman animal— lacks the capacity to understand a wrong as such, as well as the capacity to understand the significance of the compensation arrangement as being a way of restoring a relation of justice, a particular corrective response lacks part of the value that, at least in the case of human-to-human compensation, it would contain.

To better understand the worries stated above, we can disentangle different values that a given successful response to wrongdoing in the context of corrective justice typically instantiates. If *X* unjustifiably harms *Y*, a corrective response to this wrong from *X* to *Y* is successful if it instantiates at least some of the following values:

- i. The recognition of a past wrong as such.
- ii. The willingness to correct a past wrong with a fitting response.
- iii. The psychological satisfaction of the victim if the victim has the capacity to understand the normative content of the act (recognition of a wrong and the correction of it as such)
- iv. The promotion of the symbolic value of recognizing past wrongs or injustices (contributing to e.g. moral education of the wrongdoer and society more generally);
- v. Benefitting Y in some way (or at least doesn't worsen it further).

For a response to wrongdoing to constitute a successful correction of injustice, this does not need to promote all the values stated above. Take (v) as an example, namely the benefit that certain forms of compensation, like monetary compensation, normally bring about. The benefit that these forms responses to wrongdoing bring about is not present in other common responses, like apologies. Apologies are a typical instance of corrective justice mechanism. Although psychologically satisfying for the victim, apologies need not make the victim better off (in wellbeing related terms).

The values appealed to by Palmer and Cripps are reflected by (iii). As it happened with (v), (iii) isn't either a necessary value to be instantiated by a response to count as a correcting an injustice. Think of human children and some intellectually disabled human beings. These groups too lack some of the normative attitudes. Human children and some disabled human beings lack the capacity to speak out against injustices performed to them, and to voice their demands of compensation or reparation. Nonetheless despite this inability, we still believe that if an unjust harm is performed to a child, for example, this gives the child a claim of corrective justice to be for example compensated for, and thus it also gives the wrongdoer a duty to restore the fairness relation. If I missed my young child's play at school without a good reason, I believe this gives me a reason to either apologize or compensate my child for the harm caused to her, even if she is not aware that is owed such apology or compensation, and even if she cannot fully appreciate the content of my act.

Another issue worth spelling here is the issue of life-time vs post-death correction. The question becomes how discharge duties of corrective justice when the victims of an injustice are dead. It is common practice to discharge duties of corrective justice through compensation to descendants of individuals who do not longer live, but who are somehow related to the victim. Many of the animals harmed by climate change are already dead. One may suggest that while duties of corrective justice to dead humans can be discharged by compensating descendants of the victims (e.g. descendants of slaves, etc.), the same duties to non-human animals cannot be discharged in the same way because nonhuman animals lack the normative attitudes necessary to understand the value of this response.

In response to the argument above, it could be argued that non-human animals have a basic interest in other members of their species not dying prematurely or not living miserable lives. It is part of their interests that they can interact with a number of members of their species (and of other species) throughout the course of their lives. In that sense, premature deaths of individuals of their species do harm animals and goes against some of their basic interests.

To sum up this section, the idea is that there is one part of corrective justice that is not defined purely in personal terms, as in the case of other branches of justice, like distributive justice. In the case of the latter, for an individual to have an injustice-based complaint with respect to a distribution, e.g., a complaint against those who are unjustly better off than me, against those who made those better off than me, or against both groups, the individual does not need to be aware that he or she has it. So, although we may have personal duties of distributive justice (personal in the sense that these duties are owed to particular people or groups of people), these duties are not dependent on these individuals being aware that they are owed those duties.

Many would agree that we have duties of distributive justice to certain individuals, even if they are not aware that they may have been harmed or unjustly treated. Just like distributive justice has an impersonal element, so does corrective justice. If justice has intrinsic value—i.e., we value justice for the sake of justice and therefore reject injustices independently of its consequences—a world in which past injustices are recognised as such, and in which past wrongs are amended will be a better than a world in which these are never amended and therefore remained open.

There is room to come up with a full account of justice, not only distributive but also corrective, for non-human animals. I believe this is the case even if in terms of *non-ideal justice*, that is, in terms of what is feasible or possible relative to the available set of options, the account needs further specification. There is room for the existence of duties of corrective justice to non-human animals even if, as I will explain in the next section, it is unclear what particular actions can count as fitting responses of corrective justice in the case of nonhuman animals that have been wronged.

3. Fitting responses of corrective justice to nonhuman animals

The account presented here aims at showing the existence of duties of corrective justice to non-human animals, especially in the context of climate change (although they can be extended to other contexts). Climate justice is a highly non-ideal field that focuses on what is just relative to a given feasible set of options or constraints. In the case of climate change, time is one such constraint. The risks imposed by climate change are huge. This makes it urgent to come up with solutions to it. Although idealised normative theories can help in determining what duties arise in the context of climate change, they can only do so to a certain extent.

Because of the non-ideal character of climate justice, an account of corrective justice to non-human animals in the context of climate change must also be non-ideal. Such an account must not only determine the grounds for the existence of duties of corrective justice to non-human animals, but it must also specify in which ways these duties ought to be discharged given the constraints imposed both by climate change, and by the nature of the individuals the account is concerned with. This section aims at elaborating on which responses are more fitting to discharge duties of corrective justice to non-human animals given these constraints.

Let's go back to the worry above raised by Cripps and Palmer. The idea is that some standard restorative responses to wrongdoing are not fitting in the context of correction of harms to non-human animals. Take the case of verbal apologies. While perhaps some higher order cognitive animals may be able to understand the moral content of an apology, most non-human animals don't. And given that the main value that apologies instantiate is the moral reparation that it brings to the apologee, apologies are unfitting as responses that aim at correcting injustices committed to non-human animals. Apologies are also unfitting as a way of correcting wrongs committed to young infants, severely intellectually disabled, or any other marginal case in which those harmed are unable to grasp the moral character of them. But the fact that some standard corrective response to harms to most human beings are not fitting in these cases can hardly be argument against the existence of duties of corrective justice towards these individuals.³ There are ways of discharging these duties other than by performing these corrective responses.

A non-ideal account of corrective justice for non-human animals will take these constraints into consideration and will specify certain responses as more fitting than others. The fittingness of a response will be determined, for example, by whether the

³ An example of such recognition is the legal compensation of intellectually disabled individuals. A recent case of this is the joint bill to compensate those people with intellectual disabilities that were forcibly sterilized decades ago under Japan's now-defunct eugenics law.

individuals or claimants can have an interest in such responses. While the immense majority of non-human animals would not have an interest in being the recipient of an apology, they would for example have an interest in having lives that are at least worth living, if not a life with a certain amount of wellbeing, in the case of higher order animals. They could also be said to have an interest in their survival, as well as the survival of other members of their own species, as well as on the survival of members of certain other species. These interests will also determine who or which entities should be understood to be the claimant of these duties. I will come back to this point in the next section.

Given that we can recognise some of the basic interests of non-human animals, we can design fitting responses of corrective justice that take those interests into consideration and that aim at promoting them. For example, we can incorporate especial concern for reparations to non-human animals in traditional principles of corrective justice in the context of climate change, like the Just Savings Principle, the Polluters Pay Principle (PPP), the Beneficiary Pays Principle (BPP), etc. One way of doing this would be to enforce a certain sum from the liable party's payment for the correction of harms to non-human animals, after making a direct assessment of the disvalue of climate change in their lives as well future risks that it will continue to impose on them.⁴

We have reasons to suspect that corrective principles of justice that do not aim at correcting the harms to non-human animals directly (not via the correction of wrongs to humans) and therefore can often arrive to non-optimal situations for non-human animals. In some cases, they might not fulfil the discharging of corrective duties towards them and, in others, they might contribute to making animals even more worse off.

Think of legal compensation. In the legal case, compensations of wrongs committed to nonhuman animals that are enforced by the law are not necessarily directed to benefit the harmed nonhuman animal. The few legal codes that recognise nonhuman animal abuse as an offence and enforce compensation policies direct the compensation to the 'owners' of the nonhuman animal, and not the nonhuman animals themselves. If my dog bites my neighbor's dog, for example, most legal codes will require me to compensate my neighbor. An assessment of the situation will determine though that is not my neighbor who has been primarily wrong, but the dog himself. My point is that without a mechanism that ensures that a sum that is given and employed as to benefit the victim—my neighbors' dog in this case—the correction of the injustice cannot be said to have been fulfilled.

⁴ This correction could be enforced through the figure of a proxy or guardian, as it is commonly proposed in the representation of interests of future generations.

Back to the climate change scenario. Imagine that, as a result of a recommendation of some version of PPP, China and USA are forced to invest a great sum to mitigate the effects of climate change. For that they decide to invest into the implementation of technologies like Carbon Capture and Storage (CCS) and Carbon Capture and Utilisation (CCU). By funding CCS, CO_2 would be captured from sources of pollution (e.g. cement plant) and stored in some geological formation or space underground where will not be released to the atmosphere. By the implementation of CCU, CO_2 would be also captured from pollution sources, but instead of being stored, it would be employed for example to as a feedstock input to the production of oil-rich algae that will later on be used as nutritious stock-feed for farm animal production.⁵

While the fund and implementation of the procedures above can be seen as optimal forms of mitigating the effects of climate change in that they result in a more optimal concentration of CO_2 in the atmosphere, and therefore can be seen as a way of discharging the duties of corrective justice that China and USA bear due to their higher historical emissions and therefore higher contribution to the harms of climate change, we have reasons to doubt they can also discharge these duties with respect to the harms that their emissions caused to non-human animals.

The implementation of CCS means that the mass of forests that will need to be restored in terms of handling CO_2 emissions would be smaller than in pre-emissions period and still compatible with a certain amount of deforestation and degradation of green masses. Given this, eliminating a great part of the current emissions before they reach the atmosphere would be against the interests of the animals that have these green masses as their habitats and who were therefore harmed in the first place, both at the individual and at the species level.

The story is somewhat similar for CCU. Given that one of the biggest uses of CO_2 that results from this technology is nutritious stock-feed for farm animal production, it is difficult to see how this can count as a way of correcting the harms to non-human animals for past emissions. It might correct the harms to some animals, but increase the harm to others, in particular farmed animals.

To conclude, the idea is that some responses to wrongs committed to nonhuman animals are more fitting than others. Taking non-human animals' interests in mind helps designing policies that can guarantee fitting policies and principles that capable of fulfilling these duties to them.

⁵ For more details, see "Accelerating the uptake of CCS: Industrial use of captured carbon dioxide. Appendix E: CO2 for use in algae cultivation". Global CCS Institute and Parsons Brinckerhoff.

4. Claimants: individuals, collectives, species

Once we have set the case for the existence of duties of corrective justice to harmed non-human animals, it becomes crucial to identify what is the relevant individual or a relevant group of individuals that are owed correction. Candidates for claimants are species, groups of animals, individual animals, or some combination of the previous. Although predictably a difficult question, let's explore the different alternatives and challenges for each of them.

Elizabeth Cripps (2013) describes what she labels 'nature restoration approach' in the following terms:

(...) there is a lively debate going on in ecological ethics on the strengths and weaknesses of a nature restoration approach. However, it might require attemptting (if possible) to save certain ecosystems from destruction, providing alternative habitats for populations where theirs had become irreversibly hostile, and in the last instance intervening actively to preserve a species, if necessary in captivity, until sufficient habitat could be renewed. (Cripps, 107)

According to the nature restoration approach described by Cripps, the fact that some of our actions—e.g. those giving rise to climate change—have destroyed certain ecosystems which are the habitats of certain species gives us the obligation to restore them. The nature restoration approach could be justified by the following reasoning: the duty to protect nonhuman animals implies an *indirect* duty to protect certain features of ecosystems because one duty supervenes on the other. Therefore, mutatis mutandis, a duty to amend wrongs committed to nonhuman animals implies an indirect duty to restore the ecosystems in which the wronged human animals live.

But corrective duties cannot be directly owed to nature as such. I cannot consider my breaking a stone into pieces an unjust act with regards to the stone itself. Stones, ecosystems or species lack the capacity to suffer or experience wellbeing, as well as the capacity to be benefited or harmed in this welfarist sense. Furthermore, that which can improve the functioning of an ecosystem can also in many cases harm some of the individuals of the ecosystem—e.g. re-introducing predators that are at risk of extinction. (see Mosquera, 2016)

Duties to repair and to compensate are owed to individuals towards whom we can perform just or unjust acts. This means that if there is anything such as a duty to amend, restore or compensate for the destruction caused to a stone, to an ecosystem, or to certain species—of both nonhuman and human animals—this will be a derivative duty. And this derivative duty acquires its force in that it indirectly fulfils a duty to compensate or amend a wrong committed to an individual, since the destruction of an ecosystem will harm the animals who live in it.

Secondly, it would be very difficult to determine to which point an ecosystem should be restored in order for an unjust wrong to the individuals who live in it to be restituted or compensated for6. Should it be restored to the state in which it was one, two, or three centuries ago? The complexity lies in that any possible point in time that one can choose as a landmark for restoration of an ecosystem seems arbitrary in that it will be beneficial for certain species and certain animals but not for others.

Take species now. Tom Regan has defended the view that the individuals of certain species are owed compensation for being members of endangered species. In the preface to the 2004 edition of his book The Case for Animal Rights, Regan defends his nonhuman animals' rights theory from the criticism that his view does not provide the basis for an obligation to preserve endangered species. According to Regan, the critics argue that the rights view fails to do justice to our intuition that we owe something more to endangered species than we do to bountiful ones (Regan, 2004, xxxix). Interestingly, Regan responds to this criticism by saying that the rights view can accommodate the problem of endangered species by appealing to compensation:

This rights view can apply compensatory principles to animals (the East African black rhino, for example) whose numbers are in severe decline because of past wrongs (for example, poaching of ancestors and destruction of habitat). Although the remaining rhinos have no greater inherent value than the members of a more plentiful species (rabbits, say), the assistance owed to the former arguably is greater than that owed to the latter. If it is true, as it appears to be, that today's rhinos have been disadvantaged because of wrongs done to their predecessors, then, other things being equal, more should be done for rhinos, by way of compensatory justice assistance, than should be done for rabbits. (Regan, 2004: xl)

Regan understands that the subject of compensatory principles can be 'identifiable groups of nonhuman animals' (Regan, 2004: *xxxix*). We can, Regan argues, identify which groups have been subject to disadvantage in the past and compensate them for this disadvantage. Endangered species, he says, are one of these groups.

There are reasons to think that appealing to groups of individuals-rather than

⁶A similar theoretical difficulty arises regarding how to determine the optimal temperature of the planet. Given that the temperature of the planet is also under natural change, what should be the right speed by which the temperature of the planet should change—the so called 'natural temperature'; the pre-industrial-revolution-temperature of the planet.

focusing only on single individuals—is beneficial in the sense that it takes into consideration group membership and the interests of animals of a certain group or species in other animals of their species surviving or faring well.

Nonetheless, considering endangered species as potential beneficiaries of corrective or compensatory justice is problematic. If a whole group of animals, in this case a whole species, is to be the focus of compensation, Regan's proposal commits him to giving compensatory priority to every one of the individuals who are member of the endangered species—in his example, black rhinos—even in those cases in which members of other non-endangered species have been more harmed by the effects of climate change than black rhinos, but don't happen to be at extinction point.

On the other hand, taking (single) individuals as beneficiaries of compensation policies does not rule out the possibility of having of compensatory or restitutive duties towards species, too. Let us imagine that an endemic species of nonhuman animals—a species that exists only in a particular geographical region—is damaged by the spill of a chemical product that has negative effects only on this species. If every one of the individuals who are members of this species are disadvantaged because of the spill, every one of the members of this species would be owed compensation. The CEO of the company liable for the spill might be required to pay the medical treatment of all those individuals affected by the spill. Since the individuals affected by the spill are all the members of the species, this compensation would subsequently benefit the whole species in that it would, for example, prevent its extinction.

However, should preventing the extinction of this endemic species be the main aim of the compensatory policy of the company? According to what we have said the focus of the CEO should rather be those individuals who have been harmed by the split, regardless of whether the species they belong to has now become an endangered species.

There might be *derivative* duties towards species if the members of an endangered species are worse off than other individuals. Focusing on compensating species as a whole—by improving its situation, increasing the number of individuals that compose it in order to avoid extinction, etc.—can revert on compensation of the individuals that compose it. But this compensation should be a means to compensate the members of the species, not an end in itself as it was revealed in the case of natural entities like ecosystems.

Finally, although understanding claimants as individual animals would ensure that it is the most wronged individuals who get corrective measures applied to them, an individual approach like this would be highly demanding. Some have referred to this individual approach in the context of duties of justice in general and suggested that it is subject to the 'too many claimant's problem (see Cripps, 2013). Thus, such an approach, although ideally just, might be in practice highly problematic, especially given the constraints imposed by climate change.

5. Corrective vs. Distributive Justice

There is another issue that deserves further attention, and which is particularly relevant in the context of corrective justice to non-human animals, which is the issue of the baseline for correcting wrongs to non-human animals given the possibility that the lives of the animals that live in the wild are not optimal, even in the absence of climate change.

The perpetration of an injustice destroys the presumed fairness that characterises ideally fair interactions between individuals. Corrective justice aims at restoring the presumed fairness of the departing point of the individuals involved in this interaction. Take compensation as a response of corrective justice. The counterfactual perspective is present in most norms and policies that are compensatory, too. Robert Goodin (1989: 59) makes an appeal to this counterfactual perspective when discussing the interpretation that the law of torts—i.e., the one that deals with civil wrongs—gives to what should be the aim of compensation:

The aim [of compensation] is to bring him up to some baseline of well-being. That baseline to be used for reckoning the adequacy of compensation will typically be identified by reference to some status quo ante, i.e., some position that the individual himself actually enjoyed at some previous time. Thus, in the law of torts, *the baseline for compensatory damage calculations is the position that the injured party was in before the tort was committed against him*⁷. (Goodin, 1989: 59).

This counterfactual compensatory damage calculation is common in the case of wrongs committed to human beings. This is so because we *presume* that the state of affairs in which human beings are, prior to the moment in which a wrong is committed to them are, by default, not wrongful. Nonetheless, we cannot presume that the state of affairs in which most nonhuman animals find themselves is, by default, nonharmful. Most nonhuman animals find themselves in harmful situations that are contrary to their preferences which are disadvantageous for them. Some authors have for instance claimed that for the wild animals that live in the wild suffering prevails over pleasure. Among the reasons for this harmful status-quo wild nonhuman animals live in are the natural catastrophes they are exposed to, the early

⁷ Italics added for emphasis.

death of millions of new-born animals born from species with large progeny of which only a reduced number of descendants survive, and the dynamics of predation to which many of them are exposed (Horta, 2013: 113–125).

If the argument against the idyllic view of nature is sound, we have reasons to think that the counterfactual state of affairs in which animals would have not been wronged by climate change would not be beneficial to most nonhuman animals. According to this argument, it is unclear that going back to pre-industrial levels of CO2 emissions would be the optimal scenario for animals that live in the wild.

There are two ways of incorporating this concern into the account of corrective justice here proposed. First, it has been argued that we have duties of assistance to animals that live in the wild. If this is the case, and so it is the case that human beings are partly responsible for the situation of those animals that live in the wild (i.e. due to omission or having failed to fulfil our duties of assistance to those animals previously), this would be a further argument to come up with an account of corrective justice that can account for the correction of the harms of both climate change and the failure to intervene to prevent the harms incurred by wild animals.

If it is not the case that humans are (even partly) responsible for the situation of animals in the wild, perhaps corrective justice should in this case incorporate some sensitivity to distribution. An account of corrective justice that included nonhuman animals could take some distributive principle (egalitarian, prioritarian, suffiencientarian) to determine the baseline at which the correction of harms to non-human animals should be stablished.

So, although in practice distributive and corrective policies might conflate (that is, correcting injustices might turn out to also promote distributive justice), the aim of this paper is to show that there is more than the distributive justification for them. Incorporating direct concern for non-human animals in whatever principle of climate justice we use can be justified by distributive but also by corrective justice grounds.

6. Conclusion

Extending our duties of justice to non-human animals in the context of climate change proves philosophically interesting, but also practically relevant. In this paper I tried to lay the grounds for an account of corrective justice for harms committed to non-human animals, especially in the context of climate change. I have explained why certain normative attitudes are not necessary in order to count as a claimant of corrective justice. I have suggested which responses can count as fitting when it comes to discharging corrective duties of justice to non-human animals, especially in the context of climate change. I have also discussed who should be understood as the candidate for having claims of corrective justice, that is, to whom we owe duties of corrective justice, ecosystems, species, individual non-human animals, or rather, groups of non-human animals. Finally, I have raised concerned about the difficulty of determining a baseline for corrective harms to non-human animals that live in the wild, given that there are reasons to think that their situation, even after correcting for the effects of climate change, is far from idyllic. Although the basis is laid, far more needs to be said in order to be able to account and correct for all the harms and risks imposed to non-human animals in the context of climate change.

References

Bekoff M. and Pierce, J. (2009) *Wild Justice: The Moral Lives of Animals*, The University of Chicago Press.

Budolfson, M. and Spears, D. (mimeo) "Methods for quantifying animal wellbeing and estimating optimal tradeoffs against human wellbeing — and lessons for axiology, including new arguments for separability".

Cripps, E. (2013) "Harming and Protecting Non-Humans", in *Climate Change and the Moral Agent: Individual Duties in an Interdependent World*. Oxford: Oxford University Press.

Crisp, R. (2003) "Equality, priority, and compassion", Ethics, 113, pp. 745–763.

IPBES (2019) "Summary for Policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services", Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Diaz, S. et al. (eds.).

Faria, C. and Paez, E. (2020) "Why Environmentalism Cannot Beat Denialism. An Antispeciesist Approach to the Ethics of Climate Change." *In Climate Change Denial and Public Relations*, Núria Almiron, Jordi Xifra (eds.), London: Routledge, 59–73.

Global CCS Institute and Parsons Brinckerhoff (2011) "Accelerating the uptake of CCS: Industrial use of captured carbon dioxide. Appendix E: CO2 for use in algae cultivation".

Goodin, R. (1989) "Theories of compensation", *Oxford Journal of Legal Studies*, 9(1): 56–75.

Horta, O. (2013) "Zoopolis, Intervention, and The State of Nature", *Law, Ethics, and Philosophy*, 1: 113–125

IPCC (2018) "Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty" Masson-Delmotte, et al. (eds.).

McKinnon, C. (2009) "Climate Change and Corrective Justice", *Annual Review of Law and Ethics*, Vol. 17, Kant's Peace Project. Compensation, pp. 259–275.

Mosquera, J. (2016) "Are Nonhuman animals Owed Compensation for the wrongs committed to them?", in Garmendia, G. & Woodhall, A. (eds.), *Intervention or Protest: Acting for Non-Human Animals*, Vernon Press.

Palmer, C. (2016) "Living Individuals. Biocentrism in Environmental Ethics", in *The Oxford Handbook of Environmental Ethics*, Gardiner, S. M. and Thompson, A. (eds), Oxford: Oxford University Press.

Palmer, C. (2010) Animal Ethics in Context, New York: Columbia University Press.

Person, I. (1993) "A Basis for (Interspecies) Equality", Cavalieri, P. and Singer, P(eds.), *The Great Ape Project*, New York: St. Martin's Griffin, 183-193.

Regan, T. (1983/2004) The Case for Animal Rights, University of California Press.

Sunstein, C. R. and Posner, E. A. (2007) "Climate Change Justice", *The Coase-Sandor Working Paper Series in Law and Economics*, University of Chicago Law School.

De Waal, F. and Brosnan, S. F. (2003) "Monkeys Reject Unequal Pay", *Nature*, 425: 297-299.

Wsiung, W. and Sunstein, C. R. (2007) "Climate Change and Animals", *University of Pennsylvania Law Review*, Vol. 155, n.6. Symposium: Responses to Global Warming: The Law, Economics, and Science of Climate Change, pp. 1965-1740.

Göran Duus-Otterström & Edward A. Page¹

Defeating Wrongdoing: Why Victims of Unjust Harm Should Take Priority over Victims of Bad Luck²

It is sometimes suggested that victims of unjust harm should take priority over victims of other forms of harm. We explore four arguments for this view: that victims of unjust harm experience greater suffering; that prioritizing victims of unjust harm would help prevent unjust harm in the future; that it is good for perpetrators that their victims be prioritized; and that it is impersonally better that victims of unjust harm are prioritized. We argue that the first three arguments fail but that the fourth argument succeeds. Moral agents have a reason to prioritize victims of wrongdoing because this secures the impersonal value of corrective justice. However, this reason can be activated differently for different agents depending on how they are situated relative to the wrongdoing, and it may be outweighed by other factors, such as the extent of the harm that could be alleviated.

¹ Department of Political Science, University of Gothenburg & Institute for Futures Studies, goran.duus-otterstrom@pol.gu.se, Department of Politics and International Studies, Warwick University, e.a.page@warwick.ac.uk.

² The final version of this paper is forthcoming in *Social Theory & Practice*. A previous version of this paper was presented at Aarhus University. We are grateful for the comments we got during that occasion, as well as for written comments by Kasper Lippert-Rasmussen, Søren Midtgaard, and two anonymous referees. We gratefully acknowledge the financial support from Riksbankens Jubileumsfond (grant number grant number M17-0372:1).
1. Introduction

When we think about how morally pressing it is to respond to disadvantages that blight people's lives, does it matter whether this disadvantage has wrongful ("unjust harm") or non-wrongful ("unlucky harm") origins? This question has provoked a number of contrasting responses amongst moral and political philosophers working in the corrective and distributive traditions of justice theorizing.³ The question is an important one for, if unjust harm is more morally pressing than unlucky harm, we should sometimes be prepared to remedy the former instead of the latter in our normative decision making. This could potentially change how we think of a range of moral and political debates such as international aid, humanitarian intervention, environmental protection, and health policy.

In this paper, we explore four arguments that could be harnessed to show that remedying unjust harm, all other things being equal, really is more morally pressing than remedying unlucky harm. The first argument is that victims of wrongdoing suffer more than unlucky victims. The second is that assisting victims of wrongdoing is instrumental in preventing future acts of wrongdoing. The third is that it is in the interest of wrongdoers that their wrongdoing is defeated. The fourth is that it is impersonally good that wrongdoing is defeated. We argue that the first three arguments fail due to problems of contingency and moral luck. The fourth, however, succeeds. Correcting the wrongs done to victims may not always be good for the victim in the sense that this makes them better off than they were, or would have been, had the wrong not been corrected; but it is nevertheless desirable that wrongs are corrected because it ensures that the world does not carry the marks of wrong-doing. Hence, we argue, correcting wrongs is invariably a source of *impersonal* value and this value provides a reason, activated in at least some circumstances, to prioritize the alleviation of unjust harms over unlucky harms.

The paper is structured as follows. In section 2, we clarify the meaning of key terms such as "harm", "wrongdoing", "unlucky harm" and "unjust harm;" and introduce a pair of hypothetical examples that will run throughout the text. In section 3, we assess four arguments in favor of prioritizing unjust harm over unlucky harm. We note that a successful argument must point out a non-contingent value of prioritizing unjust harm and argue that the impersonal argument meets this standard. In section 4, we draw out an important implication of this argument which is relevant for the recent debates over the Beneficiary Pays Principle. In section 5, we offer some final remarks on the moral weight of defeating wrongdoing.

³ See, for example, Stemplowska (2009); McMahan (2010); Singer (2010); Tadros (2011, 105-8); Knight (2013); Parr (2016); Eggert (2018).

2. Making sense of unjust harm and unlucky harm

When exploring the claim that tending to wrongful harm is more morally pressing than tending to unlucky harm we need to clarify the concepts invoked in this claim. We take *harm* to mean a setback to someone's interests.⁴ *Unjust* harm is harm which is brought about through wrongdoing. *Unlucky* harm, by contrast, is harm which is no one's fault. For the sake of simplicity, we treat the distinction between unjust and unlucky harm as mutually exclusive and exhaustive. *Wrongdoings*, finally, are actions that are contrary to moral duty. Since our question is whether harm brought about by wrongdoing should generally take priority over harm that is no one's fault, we can be neutral between different first-order normative accounts of what makes an act wrong.

With these initial conceptual clarifications in mind, consider the following pair of examples:

Assault. Victim's daily commute takes them down the street where Scoundrel lives. Scoundrel dislikes intensely the noise that Victim's motorcycle makes. One morning, Scoundrel places a branch across the street just before Victim arrives on his bike. Victim rides into the branch and breaks a leg.

Freak Accident. Hapless travels to work every day by motorcycle. One morning, a freak gust of wind unexpectedly blows a tree branch into the street. Hapless rides into the branch and breaks a leg.

The cases *Assault* and *Freak Accident* are paradigmatic illustrations of unjust and unlucky harm. Even though both Victim and Hapless suffer the same harm, only Victim suffers unjust harm because only they are the victim of wrongdoing.⁵

We no doubt think that there are important differences between the examples that should affect how we think about helping Victim or Hapless. In particular, since Scoundrel is responsible for Victim's injury, it would be inappropriate for Scoundrel to treat the predicament of Victim and Hapless as equally morally pressing. Since

⁴ Though this is Feinberg's (1984) famous phrasing, we do not mean to take on a strongly Feinbergian approach to harm. In particular, we do not follow Feinberg in saying that an act "harms" a person only insofar as it violates an interest protected by rights (ibid. 36). Since we are not concerned with the appropriate limits of criminal law, we can afford to speak of non-rights violating harm. Note that "setting back" is ambiguous between what Tadros calls different *measures* of harm (Tadros 2016: 176).

⁵ We might think that Hapless is also a victim of wrongful harm insofar as someone neglected their moral duty to, say, prune the trees lining the street where the accident occurred. Here, however, we rely on the reader accepting our stipulation that Hapless' accident was no one's fault. Note that if Hapless' crash was the result of wrongdoing, it would be of a lesser sort than the direct attack carried out by Scoundrel. We return to degrees of wrongdoing below.

Scoundrel is responsible for the harm caused to Victim, Scoundrel has the primary duty to remedy the harm. This could, in turn, affect how third parties reflect on the situation, because if we could anticipate that Scoundrel would act on their primary duty, then others might as well assist Hapless.⁶ But let us assume that Scoundrel, now presumed dead, fled the scene immediately after his attack thereby leaving others entirely unconnected to the attack with the question of which victim to help. They may be able to help both but even in that happy circumstance they have to choose which of the two they help first thereby lengthening the suffering of the other. Does the fact that only Victim suffers unjust harm matter to the moral urgency of tending to the injuries of Victim and Hapless? That is the more difficult question.

To see the question clearly, suppose that the cost of assistance is the same and that there are no other relevant differences between Victim and Hapless: they are equally well off, financially and otherwise; they are equally within their rights to take their motorcycles to work; their physical injuries are identical; and they suffer to the same extent. Suppose also that there are no special obligations to ameliorate either person's harmful condition in virtue of some special relationship (such as friendship, family ties, or a preexisting promise of mutual aid). Does the genesis of Victim's injury *in itself* give us stronger reason to assist him? That is not clear. Since neither Victim nor Hapless could control what happened to them, perhaps we should regard their claims on our assistance as equally pressing. For example, if we could only drive one of them to the hospital, maybe we should decide by flipping a coin. Alternative decision-making strategies would not be fair, we might think, because they would recognize a normative difference in the situations of suffering agents that does not exist.⁷

3. Four ways to ground priority for unjust harm

So what could explain the common intuition that these two cases are relevantly and importantly different from the normative point of view such that bystanders (moral agents who played no role in how the wrongful harms came about) have reason to assist Victim before they help Hapless? We discuss four possibilities: that victims of unjust harm should be prioritized because this would be better *for them*, because this would *prevent* future unjust harm, because this would be better *for the perpe*-

 $^{^6}$ We disregard for the sake of argument the obvious objection that Victim would be uncomfortable with being assisted by his attacker.

⁷ This criticism has been pressed, for example, in recent critical discussions of the Beneficiary Pays Principle, typically from a luck egalitarian perspective. See, e.g., Knight (2013); Huseby (2015); Lippert-Rasmussen (2017).

trator, or because this will be better *impersonally*. We argue that only the fourth possibility succeeds.

It is important to be clear about the standard we invoke in assessing these arguments. We assume that in order for us to say that victims of unjust harm should take priority, then there must be a tight link between the fact that someone suffers unjust harm and our moral reasons to prioritize them. More specifically, the case for priority cannot depend on some contingent set of circumstances such that we occasionally have a reason to prioritize victims of unjust harm, but must instead draw on some factor which is *inherently* present in cases of unjust harm. The first three of the four arguments for giving priority to victims of unjust harm fail precisely because they fail to identify such an inherent factor.

3.1 The personal good of the victim

One reason to prioritize Victim over Hapless would be if the harm done to Victim could be shown to be greater than the harm done to Hapless. Imagine, for example, that both events occur simultaneously and in the same locale and we are first on the scene. If one is suffering more than the other then we would have a straightforward person-affecting reason to assist the worse off person. This invites the question of whether there is reason to think that the wrongfully harmed are consistently worse off than the unluckily harmed in this person affecting sense. Although Victim and Hapless share the same physical injury, it might be thought that breaking one's leg as a result of an assault causes more suffering than a blameless action or event because a victim of injustice feels additional psychological pain from knowing that their injury was brought about 'by design' (Singer 2010: 197). If that is so, then we would seem to have more reason to assist Victim since we could alleviate more harm that way.

While it is conceivable that victims of injustice do generally suffer more than victims of brute bad luck, the problem is that this somewhat contingent consideration will not translate into a general reason to prioritize victims of wrongdoing. First, any particular unlucky victim might well be psychologically more fragile than any particular victim of wrongdoing and consequently suffer more from an identical injury; and if we were to base our harm alleviation decisions solely on who is suffering more, we would then lack a general reason to prioritize the victim of wrongdoing. Second, there is a sense in which emphasizing the greater suffering of victims of wrongdoing would not quite answer the core puzzle that wrongful-versus-unlucky harm cases pose since, after the "sting" of wrongdoing is added, there would presumably be other cases to consider in which the victims of wrongdoing are evenly matched in terms of harm with victims of brute luck who suffer slightly worse injuries. If we were to look only at harm or suffering, we would have no reason to prioritize the victims of wrongdoing in these cases.

There are two problems, then, with the "good-for-the-victim" argument. First, it relies on the contingent claim that wronged victims suffer more than unlucky victims from comparable injuries. Second and relatedly, it does not provide a reason to prioritize wronged victims in those situations where unlucky victims are suffering equally to wronged victims. It is worth noting at this point that the evaluation of the good-for-the-victim argument overlooks that the normative *valence* of harm differs depending on whether it is brought about through wrongdoing or not. That is, rather than just looking at the *amount* of harm, we should also look at how *morally objection-able* the harm is. We agree with the basic intuition underlying this approach, but, as we explain below, we believe it is better developed in terms of impersonal value than in terms of the personal good of the victim. When unlucky victims and wronged victims suffer equal setbacks to their interests, we find it plausible to say that assisting them is equally pressing from the point of view of their wellbeing.

3.2 Prevention

A second reason to prioritize victims of unjust harm is prevention. The idea is that we have additional reason to respond to wrongdoing because doing so reduces the likelihood of further wrongdoing in the future. Assisting the victims of brute bad luck, by contrast, does not in itself lead to fewer unlucky victims in total. Tom Parr endorses this line of thought when he observes that defeating wrongdoing is important in the way it demonstrates that "wrongdoing will [not] be profitable" (Parr 2016: 994). Jeff McMahan (2010: 60) similarly writes that responding to unjust harm has the additional benefit that it can "deter others from acting in the same way." Such reasoning does not, of course, transfer over to unlucky harm: natural disasters and chance events cannot be deterred or rendered unprofitable by efforts to remedy their adverse human effects.

There are obviously very good reasons to try to prevent wrongdoing. What is unclear, however, is whether giving priority to the wrongfully harmed will reliably achieve this aim.⁸ While it is plausible that we can reduce future wrongdoing by threatening wrongdoers with punishment, for example, there is little reason to think that assisting victims of already *completed* wrongs would tend to have the same effect. For one thing, wrongdoers might be content with having attacked their

⁸ For this reason, it is not clear that Parr (2016) means to speak about deterrence as opposed to disincentivizing wrongdoing more generally in his discussion of why reversing wrongful transactions is important. Deterrence is a fear-based mechanism, and there is no apparent reason to think that a stated policy to reverse wrongful transactions would inspire fear. McMahan (2010: 60), by contrast, has deterrence in mind since he considers the preventive effects of military action.

victims at one point in time. If this is the case, the prospect of others tending to the victims' injuries would not dissuade them. Indeed, the incentives could conceivably go in the opposite direction: if a wrongdoer does not want his victim to suffer in a lasting way, a stated aim to assist victims of unjust harm might *increase* the likelihood of their attacking the victim.

We do not mean to suggest that remedying unjust harm can never prevent further wrongs in the future. Perhaps some wrongdoers would find it "pointless" to inflict unjust harm if they foresee that the harm will be remedied afterwards; perhaps we can prevent some victims from becoming radicalized and vengeful by making sure that they are made whole again. The point is that, like the good-for-thevictim argument, the prevention argument is prone to the objection that it justifies any priority of unjust harm over unlucky harm on an external, rather than integral, feature how this unjust harm came about.⁹ What the emerging debate on prioritizing unjust harm alleviation seeks to explore, by contrast, is whether there might be a much tighter link between being a victim of unjust harm and our moral reasons such that there is a general reason to prioritize these victims. These "tightly linked" moral reasons for giving priority to victims of wrongdoing, unlike variable factors such as the extent of victim harm or the deterrent effects of our assistance choices, will ground the case for priority in the badness of the relevant wrongdoing and so be more robust to contingencies such as those described above.

3.3 The personal good of the wrongdoer

According to the first "tightly linked" reason, prioritizing victims of wrongful harm is justified, perhaps counter-intuitively, through its positive effect on the wellbeing of the perpetrator. The idea is that wrongdoing—at least when it is intentional—can be a source of disvalue for the wrongdoer and so correcting this wrongdoing could conceivably be of benefit to the wrongdoer as well as the victim. Parr writes, for example, that wrongdoing is bad *for the wrongdoer since it* "morally defiles" her (Parr 2016: 994; see also McMahan 2010: 60; Tadros 2016: 1–2). If this is correct, one reason to assist victims of unjust harm could be that, in reversing the effects of wrongdoing, it makes this stain on the perpetrator's life go away. We should place priority on assisting Victim not because Victim is worse off than Hapless, but because this course of action, in addition to mending a broken leg, provides a personal benefit to Scoundrel that can only be captured when wrongdoing is corrected. Assisting Victim "saves Scoundrel from themselves."

One question this inevitably raises is whether we should be concerned with

⁹ See, for example, Parr (2016: 994) and McMahan (2010: 60) who both suggest the prevention rationale yields far too contingent a justification of the value of assisting victims of unjust harm.

wrongdoers' wellbeing. There are three skeptical positions we could take here. First, we might think that we have no positive reason to promote the interests of someone like Scoundrel. Second and more strongly, we might think that promoting Scoundrel's interests is, in itself, morally bad because it would make them fare better than they morally deserve.¹⁰ Third, we might think that rescuing Scoundrel from themselves is objectionable because it involves paternalism. If Scoundrel does not want to be rescued from themselves, doing so anyway would seem to benefit Scoundrel in a way they would not wish to be benefited. All three positions deny that Scoundrel's wellbeing provides a valid moral reason to prioritize Victim's suffering over Hapless'.

Let us suppose, however, that promoting Scoundrel's wellbeing has some moral value even though they chose to commit a serious attack. A further, and more complicated, question is whether acting immorally is bad for someone such that we promote the wrongdoer's wellbeing when we undo the unjust harm they have caused. This will depend, crucially, on our theory of wellbeing. Consider, first, mental-state conceptions of wellbeing such as hedonism. If we think a person's life goes better insofar as she experiences more pleasure or happiness, we have at most a contingent basis for thinking that mending Victim's leg would be good for Scoundrel. Were Scoundrel to feel good about Victim's injury, the aim of promoting Scoundrel's wellbeing would speak *against* helping Victim. The same conclusion would seem to follow if we adopt desire-fulfillment conceptions of wellbeing such as preferentialism. Suppose Scoundrel's preference was for Victim to go through prolonged physical pain as "punishment" for their noisy motorbike commute. If we promptly tend to Victim's suffering, we frustrate this preference and make Scoundrel's life go less well. According to neither conception, then, does the aim of promoting Scoundrel's wellbeing reliably translate into a reason to help Victim. However, it is possible that the good-for-the-perpetrator argument is best understood as adopting a *moralized* understanding of wellbeing. If so, even if attacking Victim is something that Scoundrel wants to do, and enjoys doing, Scoundrel's successful attack may be said to make their life go worse in adding aspects to their lives that they have reason to disvalue. One such disvaluable aspect, noted in the literature, is that people are "defiled" by committing wrongful acts.¹¹ On this view, even if Scoundrel does not *realize* that acing immorally is a source of disvalue to them, efforts to defeat Scoundrel's wrongful plan could make Scoundrel's life go better since it

¹⁰ This argument could be developed in terms of comparative or non-comparative desert (Kagan 2012).

¹¹ The idea that injustice harms the wrongdoer is, for example, a prominent theme in *The Republic* (Plato 2008). It is also presumably the idea behind Parr's (2016: 994) remark, noted above, that wrongdoing is bad for the wrongdoer because it "defiles" them. See, further, Tadros (2016: 1–2) and Parfit (2017: 402–3).

would be a less defiled life.¹² Moreover, given that the account of well-being underpinning the "defilement" idea is reconcilable with unexperienced changes in the world affecting our well-being, it can conceivably supply a perpetrator-based reason to correct wrongdoing even if the wrongdoers have disappeared or died.¹³

But how is the wrongdoer's life better when others step in to reverse the effects of their wrongdoings after the fact? This question has two elements: how the wrongs may be corrected vicariously and how this correction makes the wrongdoer's life better by being less defiled. In terms of the corrective element, it does seem tolerably clear that third parties may sometimes be able to undo *some* of the effects of wrongdoing. For example, if Scoundrel's aim was to put Victim in a state of prolonged physical pain, third parties may frustrate this aim by ensuring that Victim receives prompt medical attention. Yet, moving to the defilement element, such vicarious remedial action would not undo the wrongdoing as far as Scoundrel's involvement was concerned or achieve any reduction in their defilement.

To explain exactly why this is the case, it is useful to distinguish between the wrongdoer's "input" into a wrongful plan and the "intended consequences" of their wrongdoing. The intended consequences of wrongdoing are typically something we can defeat since we can harness mechanisms of compensation or restitution to ensure that the consequences are undone but the wrongdoer's input is complete as soon as they have completed their part in ensuring the success of the wrongful plan they set in motion. This is troubling for the perpetrator-based argument since it is the input, rather than the consequences, which plays the most significant role in a wrongdoer's defilement. The reason "input" defiles perpetrators is to do with moral luck (Nagel 1979).¹⁴ When a wrongdoer successfully completes their wrongful plan, the intended consequences of the wrongdoing are largely beyond their control. Others may or may not intervene to make sure that these consequences are not allowed to stand; and the wrongdoer typically does not know which outcome will ensue. In what way, then, can someone intervening at a later stage to correct the wrongdoer's injustice redeem the wrongdoer? If two people execute similar immoral plans, but one happens to have his plan defeated at a later stage while the

¹² One potentially troubling implication of the defilement account is that it would seem to imply that our reason to defeat wrongdoing varies in strength with the seriousness of the wrongdoing. For example, on this account it seems that we should expend more energy on redeeming Hitler for Hitler's sake than on redeeming Scoundrel for Scoundrel's sake.

¹³ For taxonomy purposes, we could understand the defilement view as a moralized version of preferentialism according to which satisfying immoral preferences is bad for us or an objective-list theory according to which committing wrongdoing is a source of disvalue for the wrongdoer (Griffin 1986; Hurka 2009).

¹⁴ Nagel writes: "[w]here a significant aspect of what someone does depends on factors beyond his control, yet we continue to treat him in that respect as an object of moral judgment, it can be called moral luck" (Nagel 1979: 59).

other does not, one wrongdoer is not, for this reason alone, plausibly seen as less defiled than the other.¹⁵ This is especially clear if we imagine that the wrongdoer has disappeared or died by the time we help his victim: if the perpetrator is permanently absent then it is unclear how correcting their wrongdoing benefits *them* since the verdict over the *moral* quality of our lives is surely final by the time we die.

What we are suggesting is that wrongdoers are not generally redeemed when the wrongful harms they bring about are corrected by others after the fact; and they are certainly not redeemed if they are now dead or so distanced to their wrongdoing that they will never reflect morally on this redemption. Things might be different if we could *prevent* someone from executing wrongful acts in the first place. Suppose we could choose between preventing Scoundrel from placing the tree branch in the street and preventing the tree branch from injuring Hapless. We might think that we should prevent the former precisely because we want to save Scoundrel from being the sort of person who has tried to assault others. But this is not the sort of case we are considering. Scoundrel has fully demonstrated his commitment to the wrongful course of action by putting the branch in the street with the aim of injuring Victim. That is why Scoundrel warrants our moral condemnation as an agent who intended to do wrong even if their wrongdoing can later be defeated (at least in part). This objection seems decisive against the argument that we should prioritize victims of wrongdoing for the wrongdoers' sake. The relevant sense of perpetrator defilement is tied to what the wrongdoer took steps to achieve and defeating the intended consequences of wrongdoing does not redeem the wrongdoer when this is done by others.

3.4 The impersonal value of correcting individual injustices

The fourth—and, we argue, superior—way of arguing that remedying unjust harms should take priority over remedying unlucky harms is to appeal the impersonal value of defeating wrongdoing. Consider again *Assault*. The difference between the situations of Victim and Hapless is that the rights of Victim were violated by a specific wrongdoer, through a specific transaction, in a way that was not the case for Hapless. Or, put slightly differently, Scoundrel's wrongful behavior disrupted relations of equality between Scoundrel and Victim—thereby failing to recognize Victim's equal moral status—in a way that just cannot be said to apply to the case of Hapless. We might think that this makes alleviating Victim's suffering, to the extent

¹⁵ For a recent argument along the same lines, see Lindstad (2020). It should be noted that there is a longstanding debate about the relationship between "defilement" and moral luck. Criminal law theorists, for example, disagree about whether actual harm matters for how much punishment an offender deserves (see, e.g., Alexander and Ferzan 2009; Morse 2010).

that this corrects the wrongful transaction imposed on them by Scoundrel, a greater priority than tending to the suffering of Hapless.

This argument might initially seem vulnerable to the same contingency objections that plagued the greater-harm-to-the-victim argument. But the claim here is not that we should prioritize victims of unjust harm because of facts about their psychology or their subjective wellbeing. The claim is rather that each and every one of us has an interest in injustices being corrected regardless of whether or not we are, or were, connected to the injustice in question. Since this sense of "interest" does not draw upon what would necessarily be good for the victims, it is appropriate to think of it in terms of impersonal value. Impersonal value may be understood in several ways (Hurka 1987: 71; Matthes 2015: 1003-5). In the present context, two understandings are most relevant. First, we might view acts, events, or states-ofaffairs as impersonally valuable if it is appropriate for *all* moral agents to take their value into consideration in their moral deliberations. This might be called the "impartial" understanding since it views impersonal value as those things that can be recognized as valuable by all irrespective of their personal perspective (Nagel 1986: 140). Second, to say that acts, events, or states-of-affairs are impersonally valuable might be understood as the claim that these things are valuable irrespective of whether they make individual lives better (Temkin 2003: 76). This might be called the "impersonal" understanding of impersonal value since it focuses on the source of the value (for whom the thing is valuable) rather than from whose perspective things are judged as being valuable.

Although the two understandings are sometimes presented as rivals (Hurka 1987: 71–2), they can be seen as complements in that they explain different ways in which moral agents have reason to incorporate a concern for impersonal value into their moral deliberations (Matthes 2015: 1004-5). Once something is viewed as bad in a way that is not reducible to the way it is bad for individuals, for example, it is a natural step to recognize the value that removing this bad would have for all moral agents, even though it would not necessarily make them better off in terms of their wellbeing. The idea is that each and every person should recognize a reason to defeat impersonal bads as a corollary of recognizing a reason, detached from their own circumstances, to promote the corresponding value to which the bad undermines. In the cases we have been looking at, this value can be seen to be *corrective justice* with its corresponding bad being corrective injustice (or wrongdoing). Defeating the impersonal bad of an uncorrected wrong can in this way be seen as a burden borne by all moral agents by virtue of them having the capacity to appreciate the badness of a wrongful (or immoral) plan continuing to stand even though they may not be connected to that plan.

The appeal to impersonal value supports the argument that defeating wrong-

doing is valuable because it corrects the wrong done to the victim in a way that all of us, including the victim and perpetrator, have reason to value. The claim that completed wrongdoing may be impersonally bad, and that acting to defeat this wrongdoing may be justified for the way in which this promotes impersonal value, has been taken up in the literature in embryonic form by Parr and McMahan. Parr, for example, states that there is value to defeating "immoral plans" because "completing immoral plans is impersonally bad" (Parr 2016: 994). McMahan's (2010: 60) similarly writes that "if immoral acts are impersonally bad events, one may prevent the occurrence of an impersonally bad event [when we prevent immoral acts]." The idea is basically that everyone has reason to value the defeat of wrongdoing and this reason flows from the impersonal badness associated with the wrongdoing leaving an imprint on the world rather than the well-being gain associated with undoing it. We might usefully think of this imprint as imposing a burden on each of us to play a role in ridding the world of uncorrected wrongs even if we are not directly involved in, and can only imagine, the particular transactions that have been disrupted.¹⁶

Defeating wrongdoing for its impersonal value can usefully be contrasted with the suggestion that unjust harms may often be experienced as more harmful than unlucky harms. The idea is not that unjust harm should be prioritized because it feels worse for the sufferer than unlucky harm; it is that unjust harm is worse in respect of the additional impersonal badness it involves. It might help to give an example. All humans will at some point in their lives experience the shocking and debilitating pain of severe toothache and it can arise wrongfully or non-wrongfully. What we are suggesting is that there is a special disvalue, or impersonal badness, in toothache that arises from wrongdoing that we, as moral agents, can appreciate even if we have never suffered ourselves from toothache with such origins. We have reason to correct the wrong, to the extent we can, not merely out of a concern to eliminate the pain of a victim but also to rid the world of a wrongful toothache—a phenomenon from which we all recoil, and comprehend as impersonally bad, even if we ourselves have never experienced it.

But what, more precisely, makes wrongdoing impersonally bad such that we have reason to prioritize their victims? We propose that a promising approach lies in the idea that unjust harms make the world go other than it ought, by contrast with how we might have preferred it had gone, in the sense that it now contains wrongdoing.¹⁷

¹⁶ As an anonymous reviewer points out, some might resist the move from recognizing the impersonal badness of unjust harm to concluding that people have a reason to act so as to defeat that harm. However, we rely on the reader accepting that there is a reason to promote value, although this reason is subject to various constraints.

¹⁷ For an alternative deployment of the idea of the "world going other than it ought" in the context of wrongful acts that created benefits for innocent parties, see Lawford-Smith (2014).

What is specifically and inherently valuable about remedying unjust harm is that it removes this bad so that the world reverts to a less defiled state. Remedying the unjust harm counteracts the wrongdoing's prior success by correcting, as far as is possible, the relevant wrongful transactions so that the world is no longer defiled by the wrongdoing in question. When agents play their appropriate role in bringing about this desirable change in the world, they extricate themselves from the regret-table situation where they have reason to correct a wrong but nonetheless fail to do so. Here, we might say, the existence of moral failure on the part of those that knowingly resist the opportunity to play an appropriate role in correcting wrongful acts amounts to a *secondary* impersonal bad. Their refusal, we might say, amounts to a refusal to recognize remedial burden that the existence of uncorrected wrongs place on all moral agents.

It may seem unusual to understand corrective justice as impersonally valuable given that, when Scoundrel wrongfully assaults Victim, Victim typically has a personal interest in Scoundrel being punished, and forced to pay compensation, to correct the injustice through which they are connected. The mechanics of the relevant remedies of corrective justice appear to have the objective of benefiting Victim, and harming Scoundrel, in the standard person-affecting sense: perpetrators correct injustices by returning victims to the condition they would have enjoyed had the injustice never happened (Ripstein 2007: 1993). Nevertheless, corrective justice can be said to have a more fundamental impersonal rationale, namely, that of "righting wrongful transactions" arising between moral agents regardless of the effects this has on the personal well-being of victims and perpetrators (Gardner 2012: 28–31). According to this view, correcting wrongful transactions may, or may not, recreate the well-being distribution that would have obtained had the wrongdoing never happened; and the duty to correct is not exclusively the wrongdoers to bear since third parties are frequently in a position to defeat elements of many correctively unjust transactions (Gardner 2012: 25-28).

The impersonal, or non-person-affecting, dimension of corrective justice can be seen most clearly with the corrective remedies of restitution and disgorgement which seek, respectively, to return objects to their rightful owners ("give back") or take objects out of the hands of non-rightful owners ("give up"). These remedies seek to correct injustices arising from wrongful transactions between agents, not to make agents party to these transactions better (or worse) off. We cannot say, then, that being subject to a remedy of corrective justice would necessarily be better for Victim (or worse for Scoundrel) in terms of their wellbeing.¹⁸ However, a corrective

¹⁸ On an objectivist account of wellbeing, we could admittedly define a person's objective good in a way that makes corrective justice necessarily in their interest. But this would re-introduce the impersonal goodness of corrective justice because the reason corrective justice is objectively good for someone

remedy will always bring about an improvement from an impersonal point of view simply because the world no longer contains an uncorrected wrongful transaction. This is not to deny that we typically do have a personal interest in corrective justice being done against those who have wronged us. But there is nothing mysterious about something being simultaneously impersonally good and good for someone in particular. Note also that by saying that wrongdoing makes the world go other than it ought, we are not denying that this cannot be true for bad brute luck as well. Our point is that immoral plans add an additional source of impersonal badness which ideally should be removed. While neither Victim nor Hapless ought to have broken their leg, Victim's injury also embodies the success of Scoundrel's immoral plan. This makes it appropriate to prioritize Victim. In doing so, we offer Victim the same person-affecting gain as Hapless but, in addition, we reduce the impersonal badness of Scoundrel's wrongdoing leaving an imprint on the world. To put it in the language of burdens, by acting this way we remove the additional burden from ourselves that we live in a world that contains a wrong that has been left to stand despite it being open to correction.

The view that the imprint of wrongdoing on the world is impersonally bad is open to the challenge that wrongdoing may make things go better. Suppose, for example, that someone assaulted and killed Hitler in the 1910s. While it wronged Hitler, it is quite conceivable that this attack would have been for the best. This possibility does not speak against what we are arguing here, however, since our claim is just that the wrong done to Hitler would, when viewed in isolation, be a source of regret. We see this clearly when we imagine that Hitler instead died from a heart attack: the impersonal badness of wrongdoing is what allows us to see how this way of preventing Hitler's future genocide would have been morally better than his dying at the hands of an unjust attack. Wrongful acts *always* make the world go worse in one respect. It follows, then, that if a remedy of corrective justice undoes a wrong, then this is necessarily an improvement to the world in one respect.

If uncorrected wrongs have impersonal disvalue, the disvalue will surely vary with the gravity of the wrong. The world has gone other than it ought to a lesser degree when the wrong is minor compared to when it is major. This means that the impersonal-value argument requires some method for grading wrongdoing.¹⁹ This would be a simple matter if the intensity of the wrongdoing merely tracked the amount of unjust harm, but we find it plausible to think of wrongness as a function of the harm caused or risked as well as the actor's culpability. This raises the difficult question of how to balance harm and culpability in overall determinations of

would presumably be that it is good *simpliciter*.

¹⁹ The same holds for the good-for-the-perpetrator argument since the wellbeing loss associated with wrongdoing varies with the gravity of the wrong.

wrongness (Ryberg 2020). We can be confident that a harm purposefully caused is more wrong than, say, a similar harm recklessly caused. But absent a principled method for balancing harm and culpability, we cannot say how much greater a reckless harm would have to be before it outstrips a purposefully inflicted harm. This problem does not matter for the basic soundness of the impersonal-value argument, however, and it presents a challenge for any view that thinks about wrongdoing in a scalar way. Nevertheless, the idea that we should prioritize victims of unjust harm because doing so removes an impersonal bad does raise the question of which kinds of unjust hams are associated with which amounts of impersonal disvalue.²⁰

4. Defeating Wrongdoing and Beneficiaries

The impersonal-value argument is powerful in part because it offers a bridge between corrective justice, which is typically seen as a "local" matter between victim and perpetrator, and moral reasons that apply to people in general. The impersonal badness of uncorrected wrongdoing, we argue, can furnish a reason *for anyone* to undo the effects of wrongdoing. However, a popular view rejects the focus on general reasons and instead maintains that *innocent beneficiaries* of a wrong have a special responsibility to redress the victims of the wrong to which both are connected. In this section, we argue that the impersonal-value argument can coexist with this "Beneficiary Pays Principle" (BPP).

According to the standard version of the BPP, "involuntary receipt of benefits stemming from injustice can, in some circumstances, give rise to rectificatory obligations to the victims of the injustice in question" (Butt 2014: 336). The impersonal-value argument sheds interesting light on this principle. Consider the following case:

Vicarious Assault. Victim frequently rides a motorcycle down the street where Recipient lives. The noise of the motorcycle scares away the customers of Recipient's outdoor café. Scoundrel, wanting to improve Recipient's life, decides to

²⁰ As an anonymous reviewer points out, an interesting case to consider is structural injustice since, on one view, the hallmark of structural injustice is that it need not involve culpable wrongdoing (Young 2011). If structural injustice involves no culpable wrongdoing, then the harms it causes presents a choice. On the one hand, we could regard them as instances of unjust harm and thus subsume them under the impersonal-value argument. The challenge of taking this route is to explain the impersonal badness of the harm even though it is non-culpably caused. On the other hand, we could conclude that harms caused by structural injustice are, in terms of priority, on a par with unlucky harm precisely because culpability is absent. It is not clear which option is to be preferred, but space prevents us from pursuing the question further here.

remove this nuisance. One morning, therefore, without telling Recipient, Scoundrel places a tree branch in the road across Recipient's street just before Victim arrives. Victim rides into the branch and breaks his leg thereby preventing them from disturbing Recipient's neighborhood for several weeks.

Let us assume that Recipient is blameless for Scoundrel's attack. Recipient did not ask for, or wanted, Scoundrel to benefit their café business in this way. However, Recipient does attract more customers because of the attack. The standard version of the BPP says that this is enough to place Recipient in a special moral relationship with Victim. For example, if Scoundrel were to flee the scene never to be seen again, Recipient "takes over" the obligation to correct Victim's loss and would fulfil this duty by relinquishing the gains to Victim until either Victim's losses are eliminated or Recipient's unjust gains are exhausted (whichever comes sooner).²¹ Recipient owes this obligation specifically to Victim, and Victim is owed it specifically by Recipient, in virtue of the unjust benefits that Recipient has gained at Victim's expense.

Critics have claimed, however, that even if there is some corrective justice-type reason for innocent beneficiaries to disgorge their benefits, this reason guides them to ameliorate the suffering of victims of injustice in general rather than Victim in particular. Huseby (2015: 219) argues that it would be unfair if beneficiaries of injustice incurred special duties to "their" victim simply because they happened to be connected through a causal chain initiated by the wrongdoer. Other critics have gone further and argued that the BPP is implausible in any formulation since it would leave victims of brute bad luck uncompensated solely because the process in which they were undeservedly harmed involved neither wrongful origins nor the production of any unjust benefits (Knight 2013: 587–8).

Responding to these fairness-based concerns about the BPP, Parr (2016: 994–5) harnesses the idea of defeating wrongdoing to argue that being the intended beneficiary of injustice gives us a stronger reason to surrender unjust gains than other types of gain. If Recipient, for example, would have received the same gains out of sheer brute luck, Recipient's reason to surrender the gains would be weaker since this would not serve the purpose of defeating Scoundrel's immoral plan.

²¹ The idea that unjust benefits should be diverted to the victims of that injustice until the victim no longer experiences unjust loss (which we might call 'compensatory disgorgement') is a feature of several recent defences of the BPP (see Butt 2014: 344; Page 2016: 91). It is worth noting, however, that the BPP need not assume that the duties of unjust beneficiaries are only, or fully, discharged once they have diverted enough of their ill-gotten gain to erase the losses of the associated victims. As we argue in the text, the precise remedy selected to shift the losses and gains of unjust enrichment should reflect a deeper commitment to defeating, as roundly as possible, the immoral plan of the perpetrators of that injustice rather than a pre-specified mechanism of shifting losses and gains amongst those party to an injustice.

However, for Parr, Recipient does not have a reason to direct the ill-gotten gains specifically to Victim since it is the disgorgement that defeats Scoundrel's plan and not disgorgement to Victim. So, on first inspection, an appeal to defeating wrongdoing saves the idea of the beneficiary *paying* but at the cost of the beneficiary paying *their victim*.

The impersonal value-argument offers a new way to navigate this debate by reinterpreting the core claim underpinning the BPP that beneficiaries of injustice have corrective duties to victims to whom they are related through a common injustice. Firstly, it holds that critics of the BPP are correct that our reasons to undo injustice in the relevant cases are agent-neutral: Recipient does not have more reason to correct Victim's loss than anyone else. Secondly, however, Recipient does have a reason, unique to them, to correct the injustice done to Victim so long as they are in the best position to defeat, in the most comprehensive manner, the injustice from which they benefited. In the language of burdens introduced above, Recipient bears the same burden as any other moral agent that lives in a world blighted by an uncorrected wrong, but this burden has a unique character in that only they, in this instance, can roundly defeat Scoundrel's wrong given the nature of Scoundrel's immoral plan. It is crucial to distinguish, here, between the normative justification of and the conditions for correcting an injustice. If there is impersonal value in correcting wrongdoing, then at the level of normative justification, Recipient is under the same reason as anyone else to defeat Scoundrel's immoral plan. But Recipient might be the person for whom that reason is activated. This is because when the wrongdoer intended to benefit the beneficiary, correcting the wrong might be something only the beneficiary can do (or do fully). If this is true, we can say that a "condition" for correcting the injustice is that the beneficiary is the one who corrects it while also maintaining that it is in the interest of all, as an impersonal value, that this wrong be corrected.²²

An example can help explain the distinction between normative justification and conditions. Suppose everyone has reason to rescue cats that are stuck in trees. Suppose further that a particular cat can only be rescued by extraordinarily tall people. Extraordinarily tall people are then the ones for whom the reason to rescue this cat is activated (because successfully rescuing this cat requires that one is extraordinarily tall). By analogy, if a condition for correcting an injustice is that the beneficiary is the one who corrects it, the general reason to correct an injustice is activated only for the beneficiary in relation to this injustice. The idea of defeating wrongdoing by reversing its effects explains how this might work. When a wrong-

 $^{^{22}}$ We are here inspired by Parfit's (1986: 143) suggestion that agent-neutral reasons may be relative to agents' capabilities to act.

doer completes an immoral plan in order to benefit someone else, it is plausible to say that the wrong is "identity-dependent," that is, the proper description of the wrong refers to particular people. For example, in *Vicarious Assault*, the wrong is "Scoundrel attacks Victim in order to benefit Recipient." This implicates Recipient in the wrong and suggests that they must play a part if Scoundrel's plan is to be defeated (Duus-Otterström 2017).

What does this entail for what Recipient should do? Considering that Scoundrel's plan was to benefit Recipient at the expense of Victim, Recipient should direct her unjust gain to Victim since this would most comprehensively defeat what Scoundrel sought to achieve. Scoundrel's plan would not be defeated, for example, merely by Recipient relinquishing their unjust enrichment to another agent (an unjust gain at Victim's expense would persist), nor would it be defeated if Victim's loss were corrected by someone else while Recipient kept possession of the unjust gains (the value of the unjust gain would have been restored to the agent from whom it came but Recipient would continue to enjoy the benefit gained at Victim's expense). Since Recipient is implicated in the wrong by virtue of being the focus of Scoundrel's intention, Recipient must take remedial action in order to defeat that intention, and this involves directing the unjust gains Recipient has acquired to restore relations of justice and equality between Victim and Recipient. This is in line with standard versions of the BPP. However, we can imagine cases where defeating Scoundrel's plan does not obviously require that Recipient transfer the unjust gain they enjoy directly to Victim. Consider a situation in which a third party steps in and compensates Victim in order to undo Victim's unjust harm. Here it is possible that defeating Scoundrel's plan does not require that Recipient transfer unjustly acquired benefits to Victim. The plan might then be defeated by Recipient giving their unjust gains to some other needy agent or perhaps to society's "general pool of resources" (Goodin 2013: 487). This action would have the effect of making some other life better than it was before, but it would still serve the purpose of defeating the immoral plan in which Recipient was implicated because it would at least cancel the benefits that Scoundrel sought to give Recipient. It would, in this way, reverse the wrongful transaction between the two parties and remove from the world an uncorrected wrong.

The impersonal value of defeating wrongdoing thus explains why intended beneficiaries of an injustice should ameliorate the situation of the victims of this injustice. It is important to stress "intended" here, because the value only applies to cases where the perpetrator sought to benefit others through wrongdoing. When people benefit from injustice in an unintended or perhaps even accidental way, it is far from clear why they would have a privileged position in defeating the injustice. Thus, our argument only applies to some versions of BPP. Note also that we do not suggest that the beneficiaries somehow inherit, or are tainted by, the perpetrators' wrongful intentions (Barry and Goodin 2014: 371-2). The idea is only that the agentneutral reason to defeat wrongdoing by reversing its effects is often (but not always) uniquely activated for the beneficiaries. When a wrongdoing is best described as an intention to attack someone in order to benefit someone else, the beneficiaries must play a part if this plan is to be defeated.²³ The rest of us can ameliorate the victims' harm but unless the beneficiaries relinquish their unjust gains, the plan will not be comprehensively defeated. Moreover, as long as the victims' losses remain uncompensated, the aim of comprehensively defeating the plan also suggests that the beneficiaries have a reason to direct the gains to the victims that were intentionally wronged as part of the plan.²⁴ This way of thinking about benefitting from injustice, or "unjust enrichment" as it is also known, may seem unnecessarily complex but it does have two useful features at odds with previous interpretations. First, the remedy will be sensitive to both the details of the immoral plan and the varying corrective capabilities of the agents connected to it. Second, it makes sense of a powerful intuition that has rarely been acknowledged in the literature on the BPP, namely, that no one truly benefits from injustice since injustice is always a loss to the world, seen impersonally, that imposes a profound burden on all agents who have the capacity to play a role in its correction.²⁵

5. Conclusion

We have argued that the impersonal badness of completed wrongdoing generates an agent-neutral reason to undo the effects of wrongdoing. This reason will be activated in different ways for different agents, but a concern to limit impersonal badness is one factor that should be considered when we are faced with the choice of responding to unjust harm or unlucky harm.

The impersonal value of correcting wrongdoing means that victims of unjust harm should be given priority over victims of unlucky harms when things are otherwise equal. But might it do more than merely "break ties"? We certainly do not suggest that it would be appropriate to give priority to victims of unjust harm in all cases. To see why this would be absurd, imagine that Hapless is badly injured while Victim has a sprained ankle. It would then be highly counter-intuitive to claim that

 $^{^{23}}$ Our account bears some resemblance to, but moves beyond, Haydar and Øverland's (2014: 356) idea that it matters whether the beneficiary was a "motivational cause" of an injustice.

 $^{^{24}}$ Thus, the aim of comprehensively defeating wrong doing can solve what one of us has called BPP's "common-source problem" (Duus-Otterström 2017).

 $^{^{25}}$ See Lu (2018: 167) who makes a useful distinction between the economic gains that beneficiaries of injustice enjoy and the moral loss to all, including beneficiaries, of a world containing injustice.

we should tend to Victim's injury first. The fact that we would correct a wrong in helping Victim is still a reason in favor of helping Victim in such circumstances, but it is handily outweighed by the unequal extent of the injuries. So the impersonal value of correcting wrongdoing supplies a *defeasible*, not decisive, reason to prioritize victims of unjust harm.

How much weight should this value have compared to other relevant considerations such as suffering, numbers affected, or cost to the remedying agents? Answering such a question is unavoidably difficult. On one approach, it will involve probing cases where correcting wrongdoing goes up against, or possibly complements, other reasons. For example, just how much worse must Hapless' injuries be for considerations of personal harm to defeat the impersonal reason we have offered to tend to Victim? Will the impersonal reason to prioritize unjust harm ever prevail in cases where a greater number of people are suffering unlucky harm? Cases must be constructed and the implications of each reason must be determined. Note that it may not be a simple matter of *weighing* reasons. The impersonal value of correcting wrongdoing plans may itself be sacrificed in some instances, not because it is outweighed by considerations of suffering or numbers affected, but because ignoring it would leave the world less desecrated in some other more important respect.

An alternative approach might be to step away from the analysis of particular cases and to ask what method agents might select to settle these cases before they arise. This would have the benefit of deriving normative conclusions from different individual standpoints while guaranteeing an impartial outlook. Peter Singer suggests that agents facing the possibility of future undeserved suffering would select a decision rule obliging rescuers to prioritize those suffering more, irrespective of the cause of this suffering. In the event that the victims suffer equally, the decision rule selected would be to toss a coin (Singer 2010: 196). We have argued, by contrast, that the undeserved suffering of the victims should be supplemented by recognizing the impersonal value of correcting the wrong done to them. This value, we have suggested, can play a valuable role in decisions as to whose disadvantage to remedy, at least when person-affecting reasons are inconclusive. If they come to realize that there is value in correcting wrongdoing, there is reason to think that agents seeking a decision procedure to settle conflicts between unjust harm and unlucky harm would not settle so quickly on a coin toss to decide who to assist even if the harm prevented is the same.

References

Alexander, Larry and Kimberly Kessler Ferzan. 2009. *Crime and Culpability: A Theory of Criminal Law*. Cambridge: Cambridge University Press.

Butt, Daniel. 2014. "A Doctrine Quite New and Altogether Untenable': Defending the Beneficiary Pays Principle," *Journal of Applied Philosophy* 31(4): 336–48.

Duus-Otterström, Göran. 2017. "Benefiting from Injustice and the Common-Source Problem," *Ethical Theory and Moral Practice* 20(5): 1067–1081.

Eggert, Linda. 2018. "Harming the Beneficiaries of Humanitarian Intervention," *Ethical theory and Moral Practice* 21(5): 1035–1050.

Feinberg, Joel. 1984. Harm to Others. Oxford: Oxford University Press.

Gardner, John. 2012. "Corrective Justice, Corrected," *Diritto & Questioni Pubbliche* 12: 9–38.

Goodin, Robert. 2013. "Disgorging the Fruits of Historical Wrongdoing," *American Political Science Review* 107(3): 478–91.

Goodin, Robert and Christian Barry. 2014. "Benefiting from the Wrongdoing of Others," *Journal of Applied Philosophy* 31(4): 363–376.

Griffin, James. 1986. *Well-Being: Its Meaning, Measurement and Moral Importance.* Oxford: Oxford University Press.

Haydar, Bashar and Gerhard Øverland. 2014. "The Normative Implications of Benefiting from Injustice," *Journal of Applied Philosophy* 31(4): 349–362.

Hurka, Thomas. 1987. "Good' and 'Good For'," Mind 96(381): 71-73.

Hurka, Thomas. 2009. Perfectionism. New York: Oxford University Press.

Huseby, Robert. 2015. "Should the Beneficiaries Pay?" *Politics, Philosophy & Economics* 14(2): 209–25.

Kagan, Shelly. 2012. The Geometry of Desert. New York: Oxford University Press.

Knight, Carl. 2013. "Benefiting from Injustice and Brute Luck," *Social Theory and Practice* 39(4): 581–98.

Lawford-Smith, Holly. 2014. "Benefitting from Failures to Address Climate Change," *Journal of Applied Philosophy* 31(4): 392–404.

Lindstad, Sigurd. 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, Kasper. 2017. "Affirmative Action, Historical Injustice, and the Concept of Beneficiaries," *Journal of Political Philosophy* 25(1): 72–90.

Lu, Catherine (2018) *Justice and Reconciliation in World Politics*. Oxford: Oxford University Press.

Matthes, Erich. 2015. "Impersonal Value, Universal Value, and the Scope of Cultural Heritage," *Ethics* 125(4): 999–1027.

McMahan, Jeff. 2010. "Humanitarian Intervention, Consent, and Proportionality," in *Ethics and Humanity*, ed. Ann Davis, Richard Keshen, and Jeff McMahan. Oxford: Oxford University Press.

Moore, Michael. 2010. *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics*. Oxford: Oxford University Press.

Nagel, Thomas. 1979. Mortal Questions. Cambridge: Cambridge University Press.

Nagel, Thomas. 1986. The View from Nowhere. New York: Oxford University Press.

Page, Edward. 2016. "Qui bono? Justice in the Distribution of the Benefits and Burdens of Avoided Deforestation', *Res Publica* 22(1): 83–97.

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

Parfit, Derek. 2017. *On What Matters: Volume Three*. Oxford. Oxford University Press.

Plato. 2008. The Republic. Benjamin Lovett transl. Digireads.

Ripstein, A. 2007. "'As if it Never Had Happened'," *William and Mary Law Review* 48(5): 1957–97.

Ryberg, Jesper. 2020. "Proportionality and the Seriousness of Crimes," in *Of One-Eyed and Toothless Miscreants*, ed. Michael Tonry. Oxford: Oxford University Press.

Singer, Peter. 2010. "Bystanders to Poverty," in *Ethics and Humanity*, ed. Ann Davis, Richard Keshen, and Jeff McMahan. Oxford: Oxford University Press.

Stemplowska, Zofia. 2011. "On the Real World Duties Imposed on us by Human Rights," *Journal of Social Philosophy* 40(4): 466–87.

Tadros, Victor. 2011. The Ends of Harm. Oxford: Oxford University Press.

Tadros, Victor. 2016. Wrongs and Crimes. Oxford: Oxford University Press.

Temkin, Larry. 2003. "Equality, Priority or What?" *Economics and Philosophy* 19(1): 61–87.

Young, Iris Marion. 2011. *Responsibility for Justice*. New York: Oxford University Press.

Henrik Andersson, Eric Brandstedt & Olle Torpman¹ Review Article: The Ethics of Population Policies

This is a review of contemporary philosophical discussions of population policies. The focus is on normative justification, and the main question is whether population policies can be ethically justified. Although few analytical philosophers have directly addressed this question - it has been discussed more in other academic fields - many arguments and considerations can be placed in the analytical philosophical discourse. This article offers a comprehensive review and analysis of ethically relevant aspects of population policies evaluated on the basis of the main ethical theories. This analysis is preceded by a brief historical contextualisation of when and how population policies became ethically contentious and how this relates to philosophical debates in environmental ethics, population ethics and political philosophy. The article also includes a conceptual analysis of population policies in which the empirical intricacies around individual fertility decisions are sorted out and the different ways in which they can be affected are categorised in a taxonomy that highlight the most relevant ethical aspects of population policies. The ethical analysis shows that while population policies can be justified on the basis of most ethical theories, it all depends on what prior assumptions are made about what is at stake.

 $^{^{1}}henrik. and erss on @fil.lu.se, eric.brandstedt @mrs.lu.se, olle.torpman @philosophy.su.se \\$

1. Introduction

Population policies are once again presented as a necessary means to reduce humanity's impact on nature and to save ourselves from ecological catastrophe. This time around it is the ever-worsening problem of climate change that is taken as a reason to 'discuss the elephant in the room', that is, how to limit world population. As always when this discussion is had, there is plenty of ethical controversies. The ethics of population policies is a topic tainted by the history of patriarchal, racist and colonial oppression it is part of. According to some critics, it is not meaningful to even try to justify population policies, they are instead best left out of political discussions altogether. But this jumps to conclusions. Recently, the topic has been explored also by philosophers (see e.g. Conly 2016; Coole 2018; Hedberg 2018; Gheaus 2019; Pinkert and Stickert 2020; and additional references discussed below) and this incipient discussion suggests that it is not clear what conclusions to draw about whether population policies can be ethically justified. This is the subject matter of this review article.

The focus is on the contemporary philosophical discussion of population policies, or more specifically on whether one can justify policies that aim to limit the size of populations. This is a normative investigation and it is the ethical justification we are interested in analysing and scrutinising. The ethical reasons for or against population policies are not always plain in sight, though, but rather often obscured in reasoning that must be reconstructed to get to the principled ground of what is at stake. This is partly due to the fact that historically, it is a topic to which few analytical philosophers have contributed to. But we contend that these non-philosophical discourses around population policies are relevant to consider in the search for their ethical status. Indeed, we believe that these larger academic and public debates on overpopulation policies form an important backdrop against which the ethics of population policies must be considered.

The article is structured in the following way. We begin in the following section with a brief historical contextualisation of population policies. Thereafter, in section three, we take a step back to consider another necessary prerequisite for a meaningful ethical analysis, that is, the empirical and conceptual intricacies around population policies. We provide a conceptual analysis of population policies that highlight the different ways in which fertility decisions and population sizes can be affected by those who so desire. The most important distinctions made are summarised in a visual taxonomy which illustrates the dimensions in which population policies should be evaluated ethically. Thereafter we turn to this evaluation and do so in a systematic way by considering whether population policies can be justified on the basis of the main ethical theories on offer, that is, ecocentric environmental ethics, consequentialism, libertarianism, feminist ethics, and theories focused on fairness. This systematic review of both the concept and ethical justification of population policies gets to the bottom of some of the ethical controversies that again has played out in recent years.

2. Historical Background

The academic discussions about population policies can be traced back to Thomas Malthus' *An Essay on the Principle of Population* (1798). Malthus argued that the human species has a natural propensity to propagate and that this stands in the way for an improvement of the well-being of the population; in particular as the supply of natural resources at best develops linearly, whereas population growth is exponential. Thus, growth of the food stock does not lead to higher levels of wellbeing, but more people and lower average wellbeing. When population growth is larger than food production growth, catastrophe looms.

It was not until the end of the 1960s, however, that a general fear spread that the population of the world was too large and that the uncurbed population growth would lead to everyone's despair.² The main source of this was Paul and Anne Ehrlich's *The Population Bomb* (1968). The Ehrlichs argued that much of the suffering in the world can be explained by overpopulation and that this raises the question of how one ought to reduce the world population. They argued that population control can take the form of incentives and penalties, but they also recognized that clearly coercive means may be needed. One example, they argued, is "the addition of temporary sterilants to water supplies or staple food. Doses of the antidote would be carefully rationed by the government to produce the desired family size" (1968, p. 130-131).³

Garrett Hardin (1968) came to a similar conclusion in arguing that coercive

² There were, of course, those who disagreed and put forward more positive views of population growth. The most famous example is Esther Boserup's *The Conditions of Agricultural Growth* (1965). Boserup argued that, as necessity is the mother of invention, population growth will lead to more efficient agricultural production. Another important criticism to this line of thought can be found in Julian Simon's *The Ultimate Resource* (1981) and *The Resourceful Earth* (1984) in which he argues that, roughly, a growing population leads to innovation, and when scarcity of a resource raises its price alternative resources will be found. This belief made him challenge Paul Erlich in a wager on the price development of metals, as he believed that their price would not rise with increasing scarcity.

³ A few years prior, in 1964, economist Kenneth Boulding proposed that a system of marketable procreation licences would meet the overpopulation problem in the most ethical way: "Each girl on approaching maturity would be presented with a certificate which will entitle its owner to have, say, 2.2 children, or whatever number would ensure a reproductive rate of one. The unit of these certificates might be the "deci-child," and accumulation of ten of these units by purchase, inheritance, or gift would permit a woman in maturity to have one legal child. We would then set up a market in these units in which the rich and the philoprogenitive would purchase them from the poor, the nuns, the maiden aunts, and so on" (Boulding 1964, p. 135).

population control is necessary for avoiding the "tragedy of the commons" – that is, in order to prevent individuals from overexploiting commonly owned or managed resources. He wrote: "A finite world can support only a finite population; therefore, population growth must eventually equal zero" (1968, p. 1243). According to Hardin, this is not a technical problem that can be solved with new technologies, nor can one appeal to people's conscience: In the long run, those who do not heed that advice will give birth to more children, many of which will have the same disposition to propagate, which means that population growth will just accelerate.

This left Hardin with coercive means as the only viable solution for curbing population growth. He does not point to any specific means which he thinks should be adopted but argues that coercion may be justified. To illustrate this, he uses the example of how we would treat a bank robber. We would not appeal to his sense of responsibility to get him to stop robbing banks. Rather, we would say that the money in the bank is not a common, and make sure that our society is not constructed in such a way that this could be perceived as a common. There are enforceable rules that prohibit bank-robbing. Similar rules would be needed with regards to procreation.⁴ Infamously, Hardin did not target everyone's reproduction equally, but held what must be described as a white nationalist or racist view about who was primarily responsible; it was the poor people in developing countries who were the cause of the problem and the 'fortunate minorities' in developed control had to impose population control there because they are all in the same lifeboat which otherwise would sink (Hardin 1968).

Around the same time, the Club of Rome released its report *The Limits to Growth* (1972), issuing stark warnings about ecological collapse due to (at least in part) overpopulation. The general anxieties around overpopulation expressed here, as well as by the Ehrlichs and Hardin, influenced the philosophical discussions as much as the public debate. The general discussions of environmental problems around the time created several new subfields within moral and political philosophy. An obvious example is *environmental ethics*, which studies the impacts of humanity on nonhuman nature and the responsibility of humans to care for the environment. Indeed, the academic field of environmental ethics arose in part due to the ecological impact of human population growth.⁵

Another outlet for discussions on population size was in *population ethics*, which took form at the end of the 1960s and beginning of the 1970s with contributions from Jan Narveson and Derek Parfit. Here, worries about overpopulation were addressed

⁴ This is not an analogy that is meant to tell us something about commons, rather it is meant to show us that coercive means can be justified.

⁵ For a more contemporary discussion similar to the *The Limits of Growth*, see e.g, J. Rockström et al. (2009).

on a much higher level of abstraction – safe from political controversies. The questions addressed were formulated in terms of how to value the future. As Katarina Forrester (2019, p. 173) notes, when the "racist and civilizational discourse of overpopulation would gradually become politically toxic for liberals and the left as antiracist critiques of eugenics, sterilization, and population control gained traction", the move towards a higher level of abstraction made population ethics durable. One could, however, argue that even if these philosophers took their views about the value of future populations to be policy-neutral, their ethical underpinnings connected to a "technocratic theory of government", which "was historically associated with colonial practices of population control and eugenics" (Forrester 2019, p. 181). Some population ethicists also drew out the policy implications of their ethical views, such as Narveson (1967) who argued that no one has the right to produce a child with a miserable life that would only burden the public.

The survivalist tendencies expressed by Hardin also influenced political philosophers. One example is Onora O'Neill (1975), who adopted Hardin's metaphor of a lifeboat ethics to address the joint threats of famine and overpopulation. She urged that the most pressing question to ask is one of survival: Given the radical shortage of resources, how can we save as many people as possible? The answer she proposed was that there was a need for both global famine prevention policies and population policies. Which specific population policies would be needed depended on the severity of the threat and they range from 'mild to draconian', from contraception to sterilization (O'Neill 1975, pp. 276f). O'Neill's problem formulation was widely shared at the time by many other liberal philosophers who focused on famine prevention and international humanitarianism. It also related to ongoing policy discussions, such as the Brandt Report (1980).

As time passed, population policies became more and more politically toxic. Eventually, it was no longer viable to relate to individual reproduction from a topdown humanitarian perspective. Individual rights and the rights of the family to procreation took centre stage. The International Conference on Population and Development (IDPD), in Cairo 1994, marked this shift in the general attitude towards population control. Principle 8 in the program of action that was agreed upon states that:

Everyone has the right to the enjoyment of the highest attainable standard of physical and mental health. States should take all appropriate measures to ensure, on a basis of equality of men and women, universal access to health-care services, including those related to reproductive health care, which includes family planning and sexual health. Reproductive healthcare programmes should provide the widest range of services without any form of coercion. All couples and individuals have the basic right to decide freely and responsibly the number and spacing of their children and to have the information, education and means to do so.

The emphasis here is on the right to decide freely and the right to not be subject to coercive means.⁶ This so-called "Cairo declaration" is part of the political backdrop against which contemporary philosophers must position themselves in arguing about population policies. The other part is the history of racist, sexist and colonial practices which have characterized the implementation of anti-natalist policies. One might wonder whether it is possible to justify population policies in this hornet's nest.

3. Conceptual and Empirical Clarifications

Before assessing population policies from an ethical point of view, certain conceptual and empirical clarifications are needed. Broadly construed, a population policy is a measure with the intention to affect the *pattern* of a population, e.g., the size, ethnographic distribution, and geographical spreading.⁷ Most often, however, population policies are understood merely as a means to affect population *size*.⁸ While population policies can be implemented as a means to stop or decrease population growth it can also be implemented as a means to increase population growth. In this paper, we focus on the former since this is what most of the relevant research has focused on.⁹ However, much of what we will say here will also be true for population policies that are introduced to increase population growth.

There are many ways to affect population size. One can, for example, influence people's *procreative decisions* – i.e., decisions about whether and when to have

⁶ In the Final Act of the International Conference of Human Rights 1968, sect II, item 16, the following claim can be found: "Parents have a basic human right to determine freely and responsibly the number and the spacing of their children." However, population growth is also seen as a hurdle for human rights provision. In other words, this shift has gradually happened. For a good overview of this development see Pizzarossa, L. B. (2018). For a discussion about the possibility that this formulation can be construed as a justification for population control see Freedman and Isaacs (1993).

⁷ As noted, population policies must typically involve an *intention* to affect the population pattern. We will however also discuss some policies that will have such an effect even if it is not intended (cf. Räikka 2001). To call any policy that will affect the population pattern, even if it is not the intended effect, seems to be too inclusive since most policies may have this effect to some degree.

⁸ This may be compared to Diana Coole's definition of *population control*. That is, "a policy regime designed to modify fertility trends through deliberate interference in reproductive behaviour, with the aim of influencing demographic outcomes." (2018 p. 4)

⁹ Pro-natalist policies are common in many parts of the world, and in some contexts, they are not believed to be as problematic as anti-natalist policies. However, as we shall discuss in section 4.4, pronatalist policies can be criticised on feminist grounds.

children or how many children to have. A variety of factors affect people's procreative decisions. For instance, different social, economic and cultural factors can be distinguished as relevant to individual procreative decisions, as well as factors such as education, religion, contraceptive use, abortion, immigration, cohabitation, age of marriage, female participation in the labour force, teenage fertility, and government programs, children as a source of labour or old age support, costs of raising children, health care improvements, gender equality, maternal and social support, and so on. Some of these factors have been shown to correlate with declining fertility rates, others with inclining fertility rates. The evidence of effects of population policies is, however, mixed (Balbo *et al* 2012).

Population policies can be characterized in different forms. Many existing characterizations draw on the distinction between *coercive* and *non-coercive* population policies. This is perhaps not surprising considering their history. The practice of compulsory sterilization, for example, is a shameful legacy of many societies and is today recognized as a horrendous abuse of human rights.¹⁰

However, the coercive/noncoercive distinction is not very illuminating (Moskowitz *et al* 1995; Steinbock 1995). Other features of population policies are also relevant. One distinction worth making in this respect is the one between *direct* and *indirect* population policies. A direct population policy targets procreation directly, such as sterilisation programs or family tax benefits. An indirect population policy targets procreation *via* some or other means, such as education programs. We will return to this when we discuss the libertarian approach to population policies in section 4.3.

Different population policies might also differ with respect to their geographical scope, which is captured by the distinction between *local* and *global* population policies. A local population policy aims to affect the population size within a certain geographical area, while a global population policy aims to affect the world's total population size. For instance, *restricted immigration* can be seen as a local population policy, since it affects the population size within a certain geographical area (typically within the borders of a nation-state). International efforts to promote qualitative education for all, on the other hand, would count as a global population policy.

Relatedly, there is a question of *whose* reproduction is targeted *within* the (local or global) area. Whether or not it was the intention of past policymakers, population

¹⁰ The most influential publication at the time, advocating sterilization, was Gosney and Popenoe (1929). The compulsory means were often motivated by arguing that the individuals subjected to these sterilizations would actually benefit from it, that is, on paternalist grounds. This, in turn, was often based on ideas of racial supremacy. For a general critique of eugenics and sterilisation programs, see Glover (1998).

policies have often targeted specific social groups. Sterilisation programs, for example, were often aimed at the mentally ill, the poor or more specifically poor women, or those belonging to certain ethnic minorities (Glover 1998). This leads to a suspicion that population policies have been disguised means of social control – or, more specifically, of separating or selecting socially desirable from socially undesirable citizens. A worry about teenage pregnancies may, for example, really be a worry about poor people multiplying and creating costs for society at large, a worry which is taken as justification for incentivising or nudging young working-class women not to procreate. We will get back to this in section 4.4 when we discuss a feminist approach to population policies.

Moreover, it should be emphasized that the *intended goal* of a population policy is potentially relevant to its justification. It can, for example, be the case that a population policy is justified as a means to reduce poverty, but not as a means to prevent biodiversity loss. As we shall see below, different ethical views yield differrent implications regarding *which* goals are relevant in this respect. This also suggests that it is important to consider whether population size should be understood as an end in itself or merely as a means to some separate end – such as alleviating poverty or preventing further biodiversity loss.

In relation to this, it is relevant to consider whether the goals, that would potentially justify a population policy, could be achieved by *other* means than population policies. Indeed, social goals can be met in different ways by targeting different factors in the complex causal web of social interaction. Although reducing fertility rates could be *one* way of reducing inequality and improving life expectancy in a society, it is not the *only* way of doing so. Another way would be to redistribute social goods. As an example, Hartmann (1995, 283) points out that successful demographic transitions in Cuba, Sri Lanka, Korea and Kerala cannot be explained in terms of population control. Instead, she argues, they are due to factors such as income and land redistribution, employment opportunities, social security, reductions in infant mortality, improvements in the position of women, and accessible health care and education.

The above shows some of the conceptual and empirical complexities around population policies. There are ethical questions in relation to all aspects of these conceptual and empirical complexities, as we will see in what follows. It may therefore be helpful to use the following graph, which illustrates the most important dimensions, to focus the attention on the ethical analysis that now follows.



In accordance with this graph, a population policy can be characterised as being anything from (i) coercive to noncoercive, (ii) means to ends-oriented, and (iii) targeting local or global demographic factors. Consequently, a population policy can at least in part be identified depending on where it is situated on this three-dimensional graph.

Moreover, this framework can be applied at an *individual* as well as a *collective* level of morality. This relates to the distinction between *personal* and *public* morality. In other words, it can be investigated to what extent ethical justification can be given for (i) governmental or non-governmental population policies to decrease the human population, and (ii) individual people's measures taken to influence others to have fewer children. While it is quite clear that collective population policies might be relevant for sustainability reasons, one might question whether an individual's choice to have fewer children is at all relevant in such regard. However, just as a collective's (e.g., a nation's) ecological impact is the product partly of the population factor, an individual's ecological impact is also the product partly of the population factor (in terms of reproduction). Murtaugh and Schlax (2009) and Wynes and Nicholas (2017) argue that by choosing to have fewer children, an individual can – other things being equal – lessen their ecological footprint compared to what it would be had they chosen to have more children (see van Basshuysen and Brandstedt 2018 for criticism). As we shall see below, however, most of our discussion will concern public population measures.

4. An Ethical Evaluation of Population Policies

In this section, we assess population policies on the basis of some influential ethical views. We start with the ecocentric and consequentialist approaches that are most permissive towards population policies in general. We then move towards more liberal approaches, including the libertarian approach and the feminist approach, that tend to be more restrictive.

4.1. The Ecocentric Approach: A Case for Reducing the Human Population

The most apparent normative defence of population policies comes from *ecocentric* moral theories. Ecocentric moral theories employ a holistic worldview according to which so-called "ecological wholes" – such as ecosystems, species, biotic communities, etc. – have *direct moral standing*. Nonhuman parts of nature have a right to exist *for their own sake*, irrespective of whether they are useful for humans. A conclusion that is often drawn from such theories is that humans have no right to infringe on these natural entities.

What is characteristic of ecocentric theories, compared to the human-centred theories discussed below, is their axiology. While human-centred moral theories typically endorse only human-related values – such as *human* well-being, autonomy, perfection, etc. – ecocentric moral theories endorse environment-related values – such as *ecosystemic* integrity, beauty, stability, biodiversity, etc. Sometimes these values are considered constituents of the "well-being" of ecosystemic wholes. Whether an act is right or wrong depends, thus, on how it affects these values – or, in other words, whether it promotes or counteracts the well-being of ecosystemic wholes (Keller 2010).

It is not entirely clear what more specific action recommendations are implied by such views, but we shall not explore the fine details of the ecocentric approach but rather its more general implications for population policies. We will, however, distinguish between *radical* and *moderate* ecocentric theories (Callicott 2013). On radical ecocentric theories, such as, e.g., Aldo Leopold's *Land ethic*, these ecocentric values are the *only* things that matter for the rightness of an action (Leopold 1949). As this implies, radical ecocentric theories allow – or even *require* – population policies to be used whenever that is needed for safeguarding ecocentric values. Such theories are highly implausible. For instance, they are "ecofascist", as Tom Regan has argued (1983), since they do not allow individual humans inviolable rights and thus open up the possibility of sacrificing individuals for the sake of ecological wholes. For that reason, moderate ecocentric theories might turn out to be more plausible. According to such theories, human beings possess direct moral standing just as ecocentric wholes do. Human-centred values are supposed to count *in addition* to the ecological values. Human well-being must thus be taken into consideration *as well as* the well-being of ecosystemic wholes. Typically, moderate ecocentric theories imply that humans have no right to use nature *over and above* what is required for satisfying basic human needs.

This notwithstanding, moderate ecocentric theories allow for quite substantial population policies too. Consider one of the most famous moderate ecocentric moral theories, *deep ecology*. The relation to population anxiety can be seen in the eight basic principles of deep ecology, formulated by Arne Naess and George Sessions. The fifth principle, for instance, states that: "The flourishing of human life and cultures is compatible with a substantial decrease of the human population. The flourishing of nonhuman life requires such a decrease." The eighth principle states that: "Those who subscribe to the foregoing points have an obligation directly or indirectly to try to implement the necessary changes." According to deep ecology, we thus have an obligation to decrease the human population (Naess and Sessions 1984).

It is not clear from the deep ecological principles *how* such a decrease in the human population should be brought about. Given the moderate ecocentric stance of taking into account human-centred values alongside ecocentric values, one possibility is that policies which frustrate basic needs of humans are impermissible. At a first glance, it might therefore seem that coercive population policies cannot be justified by deep ecology. At a closer look, however, things are more complicated. Indeed, even deep ecology regards humanity from a *holistic* point of view. This means that it is centred around the human *species* rather than on human *individuals*. Hence, "basic human needs" should be understood, not in terms of what is required for the survival and well-being of individual human beings, but rather in terms of what is required for the survival or well-being of the human species are consistent with both the death and suffering of a great number of human individuals.¹¹

Since there is no doubt that the human population is currently expanding at the cost of other species on Earth, as well as the toll it takes on many ecosystems, it seems quite clear that the current human population size is problematic from any ecocentric perspective. As this suggests, they all seem capable of justifying population policies *on the condition that the objective is to care for the well-being of ecosystemic wholes.* As a consequence, it seems quite clear that ecocentric moral theo-

¹¹ For a brief introduction to deep ecology see Brennan and Lo 2016.

ries can *both in principle and in practice* allow for substantial population policies of many kinds.

It should be noted, though, that things are a bit more complicated here as well. Humanity's impact on the Earth's ecological systems can be explained in terms of the I=P*A*T equation, where the ecological impact (I) is the product of *three* factors: the population size (P), this population's affluence measured in consumption of goods and services (A), and the technology with which these goods and services are produced (T). Consequently, the population factor is not the only factor by which ecosystemic health can be safeguarded. This same end could be reached through decreased consumption or improved technology, or some combination thereof. This in turn suggests that even if ecocentric theories could *justify* population policies, they *require* them only if there are no other alternatives available.

That being said, it is clear that population policies based on an ecocentric approach will be ends oriented. Depending on the specifics of the application, they can belong to the sphere of public as well as private morality, and be placed anywhere along the dimensions of coercive/non-coercive and global/local policies.

4.2. The Consequentialist Approach: A Case for Efficiency

Roughly, a consequentialist ethical theory can justify the implementation of a population policy if - and only if - it leads to better consequences than the implementation of any alternative (including other population policies as well as no policies at all). The consequentialist approach adopts a different axiology than the ecocentric approach. Typically, consequentialist theories are welfarist. In other words, they do not take into account any ecocentric values: Only the well-being of humans and other sentient beings matter for the rightness of an action. This restriction to the well-being of sentient beings implies a restriction on which population policies that can be justified. Still, one cannot *in principle* rule out any population policy from the perspective of consequentialism - not even coercive population policies. If a coercive policy leads to more overall well-being in the world, then, on a welfarist approach, it is justified. This is acknowledged by Räikkä (2001), who claims that coercive population policies may in some cases be preferable to noncoercive ones. Whether or not a population policy is justified, he argues, depends on the efficiency of the means it proposes and the potential harm of restricting individuals' procreative liberty by such means compared to the harm reduction thereby brought about through a smaller population size.

Robin Attfield (2015, p 129) gives another consequentialist argument for population policies. He argues that China's one child policy presents a case for coercive population policies. He reasons as follows: noncoercive means would have been insufficient for limiting the population growth in China and an uncontrolled population would have led to catastrophic consequences. So, while he acknowledges that China may have acted ethically wrong by limiting reproductive freedom, that is if they thereby failed to minimize harmful consequences, if the alternative for China would have been no population policy and thus catastrophe, then the coercive nature of the one-child policy is justified.

Although it is thus clear that any *type* of population policy can in principle be justified from a consequentialist perspective, nothing has been said so far about exactly how a population policy must be conducted more concretely in order to be justified. Few consequentialists have explicitly discussed the specifics of population policies in this respect. Philip Cafaro (2015) is an exception. He argues that we must take the severity of the possible consequences of climate change into consideration when assessing population policies. He proposes *restricted immigration* as a concrete population policy. By closing the borders of the state, he argues, the population size of a nation will be limited to the (procreation of) its existing members. This would thus count as a *local* population policy. However, Cafaro argues for this policy on the basis that immigration increases greenhouse gas emissions. More specifically, he argues that the US ought to severely limit immigration in order to become ecologically sustainable.

The ethical discussion around immigration policies is complex and it is far from clear that restricted immigration is justified on consequentialist grounds. Among other things, the effects on the would-be immigrants' well-being, the long-term consequences for the economy, and the badness of climate change must be taken into consideration before one can conclude that restricted immigration is *all things considered* a justified policy. In relation to this, it should be mentioned that fertility rates are, indeed, falling in parts of the world, much due to the abovementioned factors. This has had as a consequence that some express worries about a declining population. For example, Ben Wattenberg (2004) discusses the demographic challenges we face with falling fertility rates. Especially in Europe, this will have serious consequences that need to be addressed according to Wattenberg. Similarly, Bricker & Ibbitson (2019) stresses that important developments in the world have allowed women to have fewer children than previous generations which will lead to a decreasing world population with all the challenges that come with it.

Moreover, consequentialists, in general, tend to acknowledge that there are better alternatives to coercive population policies. Consequentialists, who claim that it is our moral obligation to make sure that procreation ends since existence entails more suffering than joy, still argue that coercive means should be avoided. The most well-known consequentialist having such a view today may be David Benatar. He argues that *if* the state were to implement coercive population policies (e.g., via legal
prohibitions of procreation), then it would have to "engage in highly intrusive policing and the invasion of privacy that that would entail", which would in effect lead to very bad consequences (2008, p 106). Obligatory abortions, for instance, would have the consequences that women would hide their pregnancy and give birth in places lacking the proper medical equipment which would, in turn, lead to much suffering. Even if Benatar is vague about which means to implement in this regard, it is clear from his (2020) that he thinks that given certain commonly accepted conditions, e.g., that we ought to combat global property, it follows that procreative freedom should be restricted.

This suggests that the consequentialist approach would *in practice* recommend noncoercive population policies before coercive ones (if it would at all recommend population policies before other means to increase welfare in the world). To find *principled* arguments in favour of non-coercive population policies over coercive, however, one must look elsewhere.

4.3. The Libertarian Approach: A Case for Incentivization

While coercive population policies can in principle be justified on both ecocentric and consequentialist theories, they are ruled out by many other ethical views. One common critique is that population policies tend to unduly restrict individual liberty. This critique can be supported on several non-consequentialist grounds, among which the libertarian moral theory is perhaps the most apparent.

On the libertarian approach, a population policy can be justified only insofar as it does not violate anyone's rights. More precisely, libertarianism condemns violations of *negative* rights, that is, individuals' rights to non-interference. The basic idea is that individuals should be free to do what they want insofar as they do not impermissibly restrict the freedom of others.

Since coercive population policies are *by definition* interfering with others' procreative freedom, they are typically hard – if not impossible – to justify on a libertarian ground. An exemption to this would be if the coercive policy is an instance of *self-defence*, and as such necessary to avoid interference that the would-be procreator otherwise would make. For sure, it is not only the would-be procreator that has rights against interference from the state, but also other people that have rights against interference from would-be procreators and their offspring. As Peter Vallentyne notes, "one has a duty to ensure that others are not disadvantaged in certain ways by the presence of one's offspring" (2002, p. 205). If procreators fail to comply with this duty, then other people have a right to defend themselves against such failures. Population policies might be one instance of such a defence. Perhaps coercive population policies could also be justified in extreme cases, even though no

one has violated or threatened anyone else's rights. Onora O'Neill, for instance, argues that coercive population policies can be justified only by the threat of major harm, such as "threats of war, famine, disease, poverty, pollution or overcrowding" (1979). However, we shall sidestep this possibility here in order to determine what noncoercive alternatives could be justified on libertarian grounds. For, it is not even clear to what extent noncoercive population policies could be so justified.

As mentioned above, individuals' fertility decisions are at least in part shaped by socio-economic and cultural factors. Hence, one potential means by which such decisions could be affected is through changes in these factors which incentivises people to have fewer children. We thus turn to the question of whether incentivizing population policies can be justified.

A closely related population policy is that of *nudging* – a notion introduced by Cass Sunstein and Richard Thaler (2009) in order to refer to the subtle ways to impact people's decision-making. The morality of nudging has lately been much discussed. Much of the discussion has focused on whether nudging is a form of manipulation (e.g., Noggle 2018 and Sunstein 2015). As such nudging appears to be coercive. For instance, a tax on childbearing might nudge or incentivize people to have fewer children, but such a tax could be argued to be coercive since it restricts people's liberty in a quite drastic way. The only libertarian justification for such a tax is that it is a means to internalize the social costs that come with childbearing. We return to this below.

Hickey et al (2016) give three criteria that must be met for incentives not to be coercive. ¹² First, there should be transparency about the goals, methods and outcomes of the implemented policy. Second, the incentives should be offered to the would-be procreators rather than government officials and families. Third, in order to avoid coercion to poor women, the incentives should be directed at "upstream" procreative behaviours, such as the use of birth control and other family planning practices.¹³ If they are correct an incentivizing population policy could reasonably be justified on libertarian grounds.

A concrete way of conducting incentivisation is through *preference adjustment*. Preference adjustment is the practice of changing the norms of individuals and the society they live in, e.g. through public campaigns (Hickey et al 2016, p. 14). The permissibility of preference adjustment depends on how it is done. If the information that causes the preference adjustment is objective and rational, then it is difficult to see what could be wrong with it. The information is then merely a catalyst for forming an informed decision. If the interventions adopt rhetorical means, how-

¹² Hickey *et al* understand an incentivizing population policy as an "attempt to influence fertility by directly altering the costs and benefits associated with certain reproductive behaviors." (p 13).

¹³ For more on the ethical dimension of incentives see e.g., Ruth W. Grant 2012.

ever, then it may be harder to justify. Although rhetorical tactics – such as emotional appeal and celebrity endorsement – would be a very efficient means of changing the preferences of a population, and much harder to justify (Ryerson 1994).

Diana Coole admits that incentives and disincentives often are coercive but argues that some incentives and disincentives can still play a part in neoliberal governance. She concludes that a case can be made for reducing the size of the population. However, two provisos must be met, human rights must be protected and "if a convincing, evidence-based case is made for its current and future benefits, and following public discussion" (Coole 2018, p. 96).

This subsection concludes that a libertarian approach can justify noncoercive population policies of an incentivizing kind. There are, however, those who object to incentivization as a means to reduce population size. Roughly, the main argument is that there is no way to know that the changes that result from incentivization are fully voluntary (Mills 1999). As Betsy Hartmann (2016, p 64) elaborates, "[f]or people who are desperately poor, there is no such thing as a free choice". Also, noncoercive incentivizing population policies will likely affect women more than men. As this suggests, incentivizing population policies may be discriminatory. This leads us to a feminist critique of population policies.

4.4. A Feminist Approach: A Case for Reproductive Rights

It can be argued that incentivizing population policies are morally problematic. Moskowitz et al, for instance, highlight that incentives for contraceptive implants are often an "instrument of class prejudice and eugenic social coercion" (1995, p 2).¹⁴ This criticism is supported by what we may call a *feminist approach* according to which both gender equality and equality, in general, must be guaranteed for population policies to be justified. This kind of justification highlights *structural* problems related to population policies, which ecocentric, consequentialist and libertarian approaches neglect.

Feminist thinkers have observed that implementation of population policies often tend to target *specific groups* of people – such as women or the poor. One possible explanation is that poor women have higher fertility rates than other women. However, the poor also have much smaller ecological footprints, so when for example climate change is discussed as a problem of overpopulation this can also reflect classist, racist and sexist attitudes through which responsibility for global problems are deflected from the affluent and poor people in developing countries are seen as the cause of their own suffering. Accordingly, a population policy must

¹⁴ For more on this see Davidson and Kalmuss (1997) and Hartmann (2016).

be designed and implemented in a way that avoids perpetuating structural discrimination. It has also been argued that population programs should give women control and encourage social changes by providing women with greater opportunities (Tangri 1976).

A specific proposal put forward in this context is a right to reproduction. This is often supported by the claim that everyone has a right to their own body and to freely form important decisions with regard to it. In this context, it is also clear that pro-natalist policies can be rejected on the same ground as anti-natalist policies. Pro-natalist policies could also be accused of treating women merely as a means and infringing their right to their own bodies. This right, it is argued, would be violated if individuals were to be manipulated to have fewer children than they otherwise would. This right to reproduction is often interpreted not only as a negative right against interference in one's decisions concerning procreation but also as a positive right, which involves a right to assistance in procreation (Brake and Millum 2018). Understood in this way, the right also involves such things as child care, income support, and health services – which are typically more important to the most marginalized in society.

One argument for the positive right to reproduction is that there are certain enabling conditions, i.e., conditions that enable individuals to freely make fertility decisions that are necessary for the reproductive right to be realized. According to Corre[^] a and Petchesky (2007), the right to reproduction comes with four enabling conditions (or "principles", as they call them): (i) bodily integrity, (ii) personhood, (iii) equality, and (iv) respect for diversity. They claim that, although the social implications of these are often ignored, "[a]ll four principles, as we interpret them, both derive from and further society's interest in empowered and politically responsible citizens, including all women" (2007, p 298).15

Sara Conly has argued that the right to reproduction can be met by having just one child (2016). This implies that a reproductive right is in principle compatible with policies limiting population size. Sure, this might not be what others have in mind when they refer to the right to procreate. This more general idea of a right to procreate, i.e., the right to one's own body, to control it and to have full autonomy over decisions relating to it, maybe compromised by a population policy introduced to limit the number of children a woman gives birth to. This line of response is also available for the suggestion that one can acknowledge that there is a right to procreate but that this right may be tradable. This view, which is found in a proposal from Boulding (1964), has recently been advanced by De la Croix & Gosseries (2009)

¹⁵ Interestingly, if these conditions are accepted, then it can be inferred that the reproductive right is violated in cases where not everyone has access to family planning, and so are forced to have more children than they actually desire.

who argue that a way to deal with over- and under-population is to introduce a scheme with tradable procreation entitlements.

Still, it could be argued that rights can be exceeded, and having more than one child will cause so much damage that it would go beyond the right to procreate.¹⁶ Conly supports this view with Amartya Sen's claim that "despite the importance of reproductive rights, if their exercise were to generate disasters such as massive misery and hunger, then we would have to question whether they deserve full protection" (1996, p 1039). In a similar vein, The European Convention on Human Rights (ECHR) states in Article 8 that a public authority may interfere with people's reproductive rights if doing so is "in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others".

Since population growth contributes to e.g. climate change, which in turn threatens these values, perhaps even a feminist approach should accept some population policies. The potential hazards of population growth are for example considered by ecofeminist Donna Haraway who formulated the slogan "Make Kin Not Babies!" in order to emphasize the role of kinmaking as an alternative to biological children (2016, p.103). This multispecies kinmaking is believed to enable a free choice to not procreate in order to reach a population size of 2-3 billions without engaging ethically problematic means (see also Clarke & Haraway 2018). More generally, given feminism focus on equality, it is likely that feminist approaches would at most support *indirect* population policies. This is in line with Tangri (1976), who argues that fertility reduction should be regarded as secondary (see also Marsden 1973). Fortunately, there is evidence that policies aimed at neutralizing gender inequalities have also reduced population growth. For instance, female education is a highly efficient means for fertility reductions, since it typically leads to smaller families (O'Neill et al. 2001, Sen 1999, and Lutz et al 2014). Also, strengthening the position for impoverished women reduces their fertility. As argued by Abadian (1996, p 1793), "by attending to fundamental freedoms for impoverished women, by enhancing women's access to and control over critical resources – their capability to achieve well-being - we not only meet welfare goals but also promote a reduction in fertility". More generally, family planning services, education, and safe methods of contraception strengthen women's reproductive autonomy and often has as a consequence that the individuals choose to have fewer children. Consequently, it is not impossible to find population policies on a feminist agenda, but they tend to be indirect, means-oriented and noncoercive.

¹⁶ See also McKibben (2013) for the view that we ought only to have one child and Overall (2012) for a discussion on procreative rights and their limits.

4.5. The Fairness Approach: The Case for Internalising the Costs of Children

The last category of arguments we will survey is focused on a comparison between parents and non-parents. The main claim, in short, is that if the decision to have a child creates costs for society at large including for non-parents, then these costs should be borne by those making the decision, i.e., the parents. The alternative, i.e., to socialise the costs, is unfair on the non-parents. If this claim can be substantiated, then some kind of population policies may be justified as a means to securing that everyone gets the fair share, they are entitled to by distributive justice.

Before we can evaluate this, it is important to note a few things about the relevant costs in question. The starting point here is that creating an individual can result in either positive or negative externalities (Casal & Williams 1995). The main positive externalities are goods and services the new individual produces and which society at large can benefit from, e.g. their work, taxes, and contributions to pension systems. The main negative externalities are the costs imposed on society and those living in it by the new individual through their consumption of scarce natural resources and production of waste.

Some have argued (e.g., Casal and Williams 1995; cf. Cripps 2015) that to the extent that having children produces negative externalities, e.g., contributes to climate change, fairness demands that these costs should be internalised to the parents - the so-called 'Parental Provision view'. Paula Casal and Andrew Williams (1995) ground their argument on a Dworkian view of egalitarian justice according to which inequalities between individuals are unjustified if they result from brute luck (e.g., natural misfortune) but justified if they result from free choice. The decision to have a child is, in relevant respects, no different from other choices an individual could make (cf. Young 2001), and so there is no reason for why others should cover its costs. To the extent that the creation of an individual reduces others' share of impersonal resources, fairness demands that the parents compensate them for that loss even though they end up worse off as a result. What concrete implications this has in terms of population policies is not fully clear, but Casal & Williams (1995) argue that subsidies to parents (e.g., child allowances and tax exemptions) should be removed and perhaps new taxes imposed. Elizabeth Cripps (2015) takes a similar line arguing that having children is unfair on non-parents because the additional costs created to make it harder and eventually even impossible for them to meet their duties of basic global and intergenerational justice. In particular, it risks placing future generations in a tragic choice situation in which either they may not have any children at all or be forced to act unjustly towards their contemporaries. We are not yet, she argues, in this situation, but morally hard choices must be made

now. We may, for example, need to introduce fines on those having children and stigmatising those having many children even though such policies would be both intrusive and aggravate inequalities.

Marcel Wissenburg (1998) seems to accept the Parental Provision view but draw a different conclusion. He argues that it leads to a paradox: on the one hand, it is necessary to reduce the world population, on the other hand, introducing population policies, which would restrict the procreative liberty of those who have not yet reproduced, is unfair and incompatible with the idea of a liberal society. Wissenburg's concern seems to be a problem of non-ideal theory. If the Parental Provision view is correct, then procreative liberty should be restricted by parents having to pay the full price of having children, but these restrictions may need to be implemented gradually so as to not frustrate anyone's existing life plans.

The fairness-based argument for restrictions on procreative liberty can, however, also be challenged in other ways. One thing is that it is far from clear what an optimal world population size is, taking into account both positive and negative externalities (Greaves 2019). Adding individuals to our world now will, other things being equal, lead to some negative externalities, but may also, for example, accelerate the development of new technology.

Another way in which the argument can be challenged is in its attribution of responsibility to the parents for the environmental impact of their grown-up children and more distant descendants (Olsaretti 2017; cf. van Basshuysen & Brandstedt 2018). Olsaretti (2017) counters the Parental Provision view on several fronts. One thing is by arguing that it assumes a static, time slice perspective on society. In the dynamic real-world situation, everyone is someone's child and the claim to internalise all externalities of children would effectively spell the end to distributive justice. In other words, the Parental Provision view is incompatible with the thought that as members of a society, there are certain things we owe one another. Furthermore, socialising the costs of children does not give benefits to parents compared to non-parents, but rather gives children their fair share – everyone is still entitled to an equal share. Olsaretti (2017) does, however, recognise that population growth can make it worse for everyone, but argues that if this is so, then this is a problem that must be explained in other ways.

A final way in which the Parental Provision view can be challenged is the claim that it cannot be implemented without undermining the social bases of self-respect for children (Heyward 2012). Even 'soft' population policies, such as removing child allowances and social campaigns against having many children, would inevitably lead to collateral damage on the children born after these are introduced and give them a worse start in life than that of previous generations.

5. Conclusion

We began this review article by highlighting the conceptual and empirical intricacies of population policies and thereafter made various distinctions which resulted in a three-dimensional taxonomy for understanding the ethically relevant dimensions of population policies. This paved the way for a deeper and more detailed assessment of their ethical status. A general implication of the results of the ethical analysis we have done is that whether or not population policies are ethically justified comes down to what fundamental assumptions are made about whose fertility decisions are targeted and for what reasons, and which consequences are taken into account in the justification. It is clear that ecocentric and consequentialist approaches can in principle allow for both direct and coercive population policies of various kinds - at least insofar as the ecosystemic well-being, or the overall welfare in the world, is thereby increased. Both libertarian and feminist approaches, however, put tougher constraints on population policies. According to the libertarian approach, coercive population policies are impermissible, but certain incentivizing policies may be allowed. The feminist approach agrees with this, but further requires that the noncoercive policies take structural justice issues into account. Finally, there are issues of fairness that must be addressed in the implementation of population policies, in particular as they tend to negatively affect those already most disadvantaged. The conclusion must be that to ethically justify population policies is very problematic.

Funding

This work was funded by the Swedish Research Council Formas (grant number 2018-02334). Henrik Andersson's work was also funded by the Swedish Research Council (grant number 2018-06698).

Acknowledgements

We presented earlier versions of this article at Lund University and the Institute for Futures Studies. We are grateful for all comments we received there.

Bibliography

Abadian, S. (1996). Women's autonomy and its impact on fertility. *World development*, *24*(12), 1793-1809.

Attfield, R. (2015). *The ethics of the global environment*. Edinburgh University Press.

Balbo, N., Billari, F. C., & Mills, M. (2012). Fertility in advanced societies: A review of research. *European Journal of Population*, 29(1), 1–38.

van Basshuysen, P., & Brandstedt, E. (2018). Comment on 'The climate mitigation gap: education and government recommendations miss the most effective individual actions'. *Environmental Research Letters*, 13, 1–3.

Benatar, D. (2008). *Better never to have been: The harm of coming into existence*. Oxford University Press.

Benatar, David (2020). Famine, Affluence, and Procreation: Peter Singer and Anti-Natalism Lite. *Ethical Theory and Moral Practice*:1–17.

Boserup, E. (2014) [1965] *The conditions of agricultural growth: The economics of agrarian change under population pressure.* Routledge.

Boulding, K., (1964) *The Meaning of the Twentieth Century* London: George Allen and Unwin Ltd,

Brake, E. & Millum, J. (2018). Parenthood and procreation. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), https://plato.stanford.edu/archives/spr2018/entries/parenthood/

Brandt, W. et al. (1980). North-South. A programme for survival - The report of the independent commission on international development issues under the chairmanship of Willy Brandt. Pan Books.

Brennan, A. and Lo, Y., (2016) "Environmental Ethics", *The Stanford Encyclopedia* of *Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/ethics-environmental/>.

Bricker, D., & Ibbitson, J. (2019). *Empty planet: the shock of global population decline*. Hachette UK.

Cafaro, P. (2012). Climate ethics and population policy. *Wiley Interdisciplinary Reviews: Climate Change*, *3*(1), 45–61.

Cafaro, P. (2015). *How many is too many?: The progressive argument for reducing immigration into the United States*. University of Chicago Press.

Casal, P., & Williams, A. (1995). Rights, equality and procreation. *Analyse & Kritik*, 17, 93–116.

Callicott, J. B., (2013), *Thinking Like a Planet: The Land Ethic and the Earth Ethic*, New York: Oxford University Press.

Clarke, A. E. & Haraway, D. (eds) (2018) *Making Kin Not Population*, Prickly Paradigm Press.

Coole, D. (2018). Should we control world population?. John Wiley & Sons.

Conly, S. (2016) *One child: Do we have a right to more?*. Oxford University Press, USA.

Correa, S., & Petchesky, R. (2007). Reproductive and sexual rights: a feminist perspective. In *Culture, Society and Sexuality* (pp. 314-332). Routledge.

Cripps, E. (2015). Climate change, population, and justice: Hard choices to avoid tragic choices. *Global Justice: Theory, Practice, Rhetoric*, 8(2), 1–22.

Davidson, A. R., & Kalmuss, D. (1997). Topics for our times: Norplant coercion--an overstated threat. *American journal of public health*, *87*(4), 550–551.

De la Croix, D., & Gosseries, A. (2009). Population policy through tradable procreation entitlements. *International Economic Review*, *50*(2), 507–542.

Ehrlich, P. R. & Ehrlich, A. (1968). *The population bomb*. New York, Ballantine Books

Forrester, K. (2019). *In the shadow of justice: Postwar liberalism and the remaking of political philosophy*. Princeton University Press.

Freedman, Lynn P., and Stephen L. Isaacs. "Human rights and reproductive choice." *Studies in family planning* (1993): 18-30

Ghaeus, Anca, 2019. 'More Co-Parents, Fewer Children: Multiparenting and Sustainable Population. *Essays in Philosophy* 20(1):2

Glover, J.. 1998. Eugenics: Some Lessons from the Nazi Experience. In *The Future* of *Human Reproduction: Ethics, Choice and Regulation* Eds John Harris, SØren Holm OUP

Gosney, E. S., & Popenoe, P. (1929). *Sterilization for Human Betterment. A Summary of Results of 6,000 Operations in California, 1909-1929.*

Grant, R. W. (2014). *Strings attached: Untangling the ethics of incentives*. Princeton University Press.

Greaves, H. (2019). Climate change and optimum population. *The Monist*, 102(1), 42–65.

Haraway, D. J. (2016). *Staying with the trouble: Making kin in the Chthulucene*. Duke University Press.

Hardin, G. (1968). The tragedy of the commons. Science, 162(3859), 1243-1248.

Hartmann, B. (1995). *Reproductive rights and wrongs: the global politics of population control*. South End Press.

Hedberg, Trevor (2018). Climate Change, Moral Integrity, and the Obligation to Reduce Individual Greenhouse Gas Emissions. *Ethics, Policy, and Environment* 21(1): 64–80.

Heyward, C. (2012). A growing problem? Dealing with population increases in climate justice. *Ethical Perspectives*, 19(4), 703–732.

Hickey, C., Rieder, T. N., & Earl, J. (2016). Population engineering and the fight against climate change. *Social Theory and Practice*, 845–870.

Kaya, Y., & Yokobori, K. (Eds.). (1997). *Environment, energy, and economy: strategies for sustainability*. Tokyo: United Nations University Press.

Keller, David, R. (2010), *Environmental Ethics: The Big Questions*, Whiley-Blackwell.

Leopold, A. (1949), A Sand County Almanac, Oxford University Press.

Lutz, W., W.P. Butz, and K.C. Samir, eds. (2014), World Population and Human Capital in the Twenty-First Century (Oxford: Oxford University Press).

Malthus, Thomas Robert, Don. *Malthus:'An Essay on the Principle of Population'*. Cambridge University Press, 1992.

Marsden, L. R. (1973). Human rights and population growth – A feminist perspective. *International Journal of Health Services*, *3*(4), 567–575.

McKibben, B. (2013). *Maybe one: A personal and environmental argument for single child families*. Simon and Schuster.

Meadows, D. H., Meadows, D.L., Randers, J., Behrens, W., (1972). *The Limits to Growth; A Report for the Club of Rome's Project on the Predicament of Mankind*. New York: Universe Books.

Mills, C. (1999). The ethics of reproductive control. *The Philosophical Forum*, 30(1), 43–57.

Moskowitz, E. H., Jennings, B., & Callahan, D. (1995). Long-acting contraceptives ethical guidance for policymakers and health care providers. *The Hastings Center Report*, *25*(1), 1–8.

Naess A. & Sessions G. (1984). The basic principles of deep ecology.

Narveson, J. "Utilitarianism and new generations." Mind 76.301 (1967): 62-72.

Olsaretti, S. (2017). Children as negative externalities? *Politics, Philosophy & Economics*, 16(2), 152–173.

O'Neill, Brian C., F. Landis MacKellar, and Wolfgang Lutz. *Population and climate change*. Cambridge University Press, 2005.

O'Neill, Onora. "Lifeboat Earth." *Philosophy & Public Affairs*, vol. 4, no. 3, 1975, pp. 273–292

O'Neill, Onora, 1979, "Begetting, Bearing, and Rearing," in O'Neill and Ruddick 1979, pp. 25–38.

Pinkert, Felix and Martin Sticker, 2020. 'Procreation, Footprint and Responsibility for Climate Change'. *The Journal of Ethics*, 1–29.

Pizzarossa, L. B. (2018). Here to stay: The evolution of sexual and reproductive health and rights in international human rights law. *Laws*, *7*(3), 1–17.

Regan, Tom, 1983, The Case for Animal Rights, Berkeley: University of California Press.

Rockström, J., et al (2009). Planetary boundaries: exploring the safe operating space for humanity. *Ecology and society*, *14*(2).

Ryerson, W. N. (1994). Population communications international: Its role in family planning soap operas. *Population and Environment*, 255–264.

Räikkä, J. (2001). Coercive population policies, procreative freedom, and morality. *Philosophy & Geography*, 4(1), 67–77.

Sen, A. (2001). Development as freedom. Oxford Paperbacks.

Simon, J. (1981). The Ultimate Resource. Princeton: Princeton University Press.

Simon, J. L., & Kahn, H. eds. (1984). *Resourceful earth: a response to global 2000*. New York, Basil Blackwell.

Steinbock, B. (1995). Coercion and long-term contraceptives. *The Hastings Center Report*, *25*(1), 19–22.

Tangri, S. S. (1976). A feminist perspective on some ethical issues in population programs. *Signs: journal of women in culture and society*, *1*(4), 895–904.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

United Nations. Final Act of the International Conference of Human Rights 1968 sect II item 16.

Vallentyne, Peter. "Equality and the Duties of Procreators." *Philosophy publications*. Oxford University Press, 2002.

Wattenberg,, B. J., (2004), Fewer: *How the New Demography of Depopulation Will Shape Our Future* Chicago: Ivan R. Dee, 2004.

Wertheimer, A. (1987). Coercion. Princeton University Press.

Wissenburg, M. (1998). The rapid reproducers paradox: Population control and individual procreative rights. *Environmental Politics*, *7*(2), 78–99.

Wynes, S., & Nicholas, K. A. (2017). The climate mitigation gap: education and government recommendations miss the most effective individual actions. *Environmental Research Letters*, *12*(7), 074024.

Young, T. (2001). Overconsumption and procreation: Are they morally equivalent? *Journal of Applied Philosophy*, 18(2), 183–192.

Partha Dasgupta¹ and S.J. Beard² Optimum Population and Environmental Constraints – A Utilitarian Perspective³

In the Utilitarian tradition of Henry Sidgwick, population ethics is to be considered through the expectation of the sum of utilities of all who are ever born. This work is unsuited for application to the world we have come to know for at least two reasons. Firstly, it takes as its subject the choice to be made by an "objective social planner", to whom the interests of everyone are equivalent. Secondly, it takes as its object an entirely idealized notion of well-being that is neither derived from any particular source nor bounded by any particular constraint. In reality, our ethical theories should extend into realms in which decision-makers are not objective planners, but subjective agents, and where well-being has to be derived from consuming the resources of a finite planet. In this paper, we seek to amend this utilitarian model by altering the way individual well-beings are aggregated by decision-makers so that they reflect the conditions that are faced by those whose decisions have the greatest impact on demographic change, namely parents and prospective parents. Next, we put this theory to work on a more realistic notion of well-being that takes account of the

¹ Department of Economics and Centre for the Study of Existential Risk, University of Cambridge, partha.dasgupta@econ.cam.ac.uk

² Centre for the Study of Existential Risk, University of Cambridge, sjb316@cam.ac.uk

³ We are most grateful to Aisha Dasgupta for the many discussions we have had on the populationpoverty-environment nexus. This projectwas made possible through the support of a grant from Templeton World Charity Foundation. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of Templeton World Charity Foundation.

biosphere's ability to supply humanity with goods and services. Finally, we begin what appears to us to be the crucial next step in applying utilitarian reasoning to global population ethics by addressing some of the issues that may explain differential fertility rates around the world, including different approaches to property rights to natural resources, different ethical assumptions about the value of children, and the unequal distribution of resources. However, we conclude that there are reasons not simply to let these differences stand, but rather to use ethical insights from our model and from elsewhere to advocate for changes that are likely to be in the long term interest of humanity.

Introduction

Whilst a variety of approaches to the study of population ethics have been proposed in various disciplines, by far the greatest body of work in recent years has been in the Utilitarian tradition of Henry Sidgwick (1907), who saw the ground of binding reason for choices facing us to be the expectation of the sum of utilities of all who are ever born. We owe a great debt to this tradition and its numerous insights; however, this work is unsuited for application to the world we have come to know, for at least two reasons. The first of these is that the tradition takes as its subject the choice to be made by an "objective social planner" (Sidgwick referred to the figure as the 'Point of View of the Universe'), to whom the interests of everyone are equivalent. The second is that it takes as its object an entirely idealized notion of well-being that is neither derived from any particular source nor bounded by any particular constraint. In reality our ethical theories should readily extend into realms in which decisionmakers (hereafter DMs) are not objective planners, but subjective agents and where well-being has to be derived from consuming the resources of a finite planet.

In this paper we seek to amend this utilitarian model in two ways. First, we alter the way individual well-beings are aggregated by the DM so that they reflect the conditions that are faced by those whose decisions have the greatest impact on demographic change, namely, parents and prospective parents. We contend that the correct choice for parents cannot simply be 'derived' from what population ethics might prescribe for an ideal planner, and instead argue for an alternative theory that we call Generation-Relative Utilitarianism. Secondly, we put this theory to work on a more realistic notion of well-being that takes account of the biosphere's ability to supply humanity with goods and services. That said, our estimates are even cruder than usual in the social sciences; however, we believe that this is largely because the questions we address here have had no airing in the literature.

In Part I we review Sidgwick's Utilitarianism, outline our alternative Generation-Relative Utilitarianism and argue that this is better suited to contemporary population ethics. In Part II, we first offer a brief review of what is currently known about the state of the biosphere and what this implies about the flow of goods and services the biosphere can supply humanity on a sustainable basis. We then put Generation-Relative Utilitarianism to work to construct estimates of the optimum global population and the optimum standard of living in the face of finite biosphere. In Part III we begin what appears to us to be the crucial next step in applying utilitarian reasoning to global population ethics by addressing some of the issues that may explain differential fertility rates around the world. We begin with an analysis of how different approaches to property rights to natural resources affect the optimum population size for communities who are served by these resources; when then briefly consider other possible differentiating factors that explain the relationship between unequal global wealth and different rates of fertility. However, in the end we conclude that there are reasons not simply to let these differences stand, but rather to use ethical insights from our model and from elsewhere to advocate for normative changes that are likely to be in the long term interest of humanity.

Throughout this paper we assume that labour in conjunction with a combined measure of produced capital and the biosphere can produce consumption goods. This generalizes the corresponding exercise in Dasgupta and Dasgupta (2017), where it was assumed for simplicity that the biosphere directly produces consumption goods, like manna from Heaven.

Part I. Parental Utilitarianism and Optimum Household Decision-Making

1. Classical Utilitarian Ethics

In his statement of Utilitarianism, Sidgwick (1907: 415-416) wrote:

... if we take Utilitarianism to prescribe, as the ultimate end of action, happiness as a whole, and not any individual's happiness, unless considered as an element of the whole, it would follow that, if the additional population enjoy on the whole positive happiness, we ought to weigh the amount of happiness gained by the extra number against the amount lost by the remainder. So that, strictly conceived, the point up to which, on Utilitarian principles, population ought to be encouraged to increase, is not that at which average happiness is the greatest possible ... but that at which the product formed by multiplying the number of persons living into the amount of average happiness reaches its maximum.

This implies that the basis for evaluation is not gains and losses to people, but gains and losses in total utility. An ethics grounded on this view sees all lives as intrinsically valuable, and hence holds that the better the life is for the person, measured in terms of happiness (or utility), the greater its value. Sidgwick's utilitarianism thus asks us to evaluate alternative states of affair in terms of the sum of personal utilities where state of affairs *X* is superior to state of affairs *Y* if total utility in *X* exceeds total utility in *Y*.

This view, which Rawls (1972) called Classical Utilitarianism, has long been known to favour large populations. Applying the theory to economic models, Dasgupta (1969)

showed that it commends a standard of living for the average person that is not much higher than that at which their utility is zero (that is, life is neither good nor not good; what Sidgwick called a "neutral life"). Much work in contemporary population ethics has focused on this implication of the theory, and in particular its purely theoretical implications for population axiology (Parfit 1984, Broome 2004). However, such debates have offered little, if anything, of value to practical considerations.

2. Generation Relative Utilitarian Ethics

In developing alternatives to Classical Utilitarianism, it is worth noting that it involves two related notions: (i) Personal Happiness (or utility) as the valuable property of lives. (ii) Summation as the required operation for combining individual utilities.

Sidgwick (1907: 119–150) devotes three chapters to the statement of what he means by Personal Happiness, which he terms empirical hedonism, and thus uses this term in a way that is considerably more considered than is suggested in the frequent criticism that Classical Utilitarianism views humans to be mere pleasure machines. Nevertheless, we are reluctant to give the impression that the ethical theory we appeal to in this essay relies on Sidgwick's notion of the happiness. We shall therefore modify (i) by using the term "well-being", which is generally recognised as capturing the broadest possible conception of personal good. Griffin (1986) provides a measured, book-length analysis of the concept in its many guises, but he also develops his preferred interpretation. Briefly, he thinks of personal well-being as a measure of the extent to which one's informed desires are realized. While composing this essay we have kept Griffin's conception in mind, but the mathematical structure of the ethics that is constructed here is not tied to any formulation of well-being.

The second amendment we make to Sidgwick's theory is a substantial modification of condition (ii). When constructing his version of Utilitarianism Sidgwick took the position of an entirely objective observer, or planner, to whom "the good of any one individual is of no more importance ... than the good of any other" and which is described as the 'Point of View of the Universe.'⁴ We do not believe that this is the appropriate position from which to consider population ethics because, in reality, no such external point of view is available and the decisions that most affect demographic change are the highly subjective choices of individual parents, and prospective parents. This leads us to modify Sidgwick's theory in two key ways. Firstly, we frame all decisions at the household level, where people are choosing the impact of having children merely upon the well-being of family members, taking

⁴ Sidgwick who was in many ways a very humble man, had the grace to write "the point of view (if I may say so) of the Universe". However, posterity has preferred this shortened phrase.

account of considerations such as the availability of resources to the family and how these will need to be distributed amongst its members. Secondly, for reasons we will describe below, it leads us to aggregate individual well-beings in a way that gives less weight to the well-being of merely potential lives than actual lives. We call this view Generation-Relative Utilitarianism.

Our justification for this modified theory is that population ethics is not only merely about identifying desirable demographic states of affairs, but also about the reproductive choices people can justify to themselves. Answers to the latter problem do not follow from resolutions of the former exercise. We confirm presently that Generation-Relative Utilitarianism commends smaller populations (and hence higher living standards) than Classical Utilitarianism.⁵

The modern philosophical source that comes closest to describing this problem is Scheffler (1982), who pointed to agent-centred concerns that people, when deliberating over courses of action open to them, can justifiably use as prerogatives over agent-neutral demands (p. 20):

... a plausible agent-centred prerogative would allow each agent to assign a certain proportionately greater weight to his own interests than to the interests of other people. It would then allow the agent to promote the non-optimal amount outcome of his own choosing, provided only that the degree of its inferiority to each of the superior outcomes he could instead promote in no case exceeded, by more than the specified proportion, the degree of sacrifice necessary for him to promote the superior outcome.

In a society of, say, *P* persons, those prerogatives would apply reciprocally, meaning that the state of affairs that would ensue would be the outcome of *P* choices, each having been guided by agent-relative concerns. In population axiology the force of those prerogatives works unidirectionally. So, we make use of an attenuated version of this idea and assume that DM evaluates states of affairs on the basis of a weighted sum of personal well-beings, where the weight she places on potential well-beings of children is less than the weight they place on their own well-being. However, we also assume that the DM will know, in advance, that once this potential person becomes an actual person then they will want to share resources with the children they produce on an equal basis with themselves. There is thus a gap between ex ante and ex post concern for future children that is like nothing we can find in Classical Utilitarianism.

The failure to account for agent-centred prerogatives is not the only thing that

⁵ A more complete justification for Generation-Relative Utilitarianism is in Dasgupta (2018).

has held population ethicists back from making useful contributions to many demographic debates. Of equal importance is their failure to put theories to work in a model where well-being must be derived from a finite stock of natural capital. We turn to this in the next section.

Part II. Optimum Population in the Real World

3. Global Ecological Footprint

An enormous literature in the environmental sciences records substantial declines in global biodiversity over the past few decades and reductions in the productivity of ecosystems that have come allied to them. Those declines can be traced to the environmental and reproductive externalities people impose on one another. By "externalities" we mean the unaccounted for consequences that our actions have for others. In this paper we frame the idea of optimum population in the context of a finite biosphere. So we focus on the adverse externalities that come allied to our actions. To be sure, the externality we each impose on others is negligible; but when they are summed over us all, the consequences are not negligible. Today, growth in atmospheric carbon concentration is the canonical expression of adverse externalities, but humanity faces wider and deeper threats to our future from the species extinctions now taking place, which are also morally more reprehensible. Proximate causes of extinctions are destruction and fragmentation of natural habitats and over-exploitation of biological communities residing there. We are converting land into farms and plantations, destroying forests for timber and minerals, applying pesticides and fertilizers so as to intensify agriculture, introducing foreign species into native habitats, and using the biosphere as a sink for our waste. And they are taking place at scales that are orders of magnitude greater than they were even 250 years ago. Millennium Ecosystem Assessment (2005) reported that 15 of the 24 ecosystems the authors had reviewed world-wide were either degraded or are being exploited at unsustainable rates. Current extinction rates of species in various orders have been estimated to be 10-1,000 times higher than their average rate (about 1 per million species per year) over the past several million years (Sodhi, Brook, and Bradshaw, 2009). World Wildlife Fund (2018), for example, records that over the past 40 years there has on average been a 60 % decline in the populations of mammals, birds, fish, reptiles, and amphibians, mostly centered in the tropics. The publication points to the growth in palm oil and soya plantations, and to the construction of dams, mines, and roads as proximate causes. There is no getting away from the scale of the contemporary human enterprise.

Studying biogeochemical signatures of the past 11,000 years, Waters et al. (2016) tracked the human-induced evolution of soil nitrogen and phosphorus inventories in sediments and ice. The authors reported that the now-famous figure of the hockey stick that characterises time series of carbon concentration in the atmosphere are also displayed by time series of a broad class of global geochemical signatures. They display a flat trend over millennia until some 250 years ago, when they begin a slow increase which continues until the middle of the 20th Century, when they show a sharp and continuing rise. Despite the uncertainties, these findings put the scale of humanity's presence on the Earth system in perspective and explain why our current times have been recognized as the start of a new epoch, the Anthropocene. Waters et al. (2016) proposed that mid-20th Century should be regarded as the time we entered the Anthropocene.⁶

This is consistent with macroeconomic statistics. World population in 1950 was about 2.5 billion and global output of final goods and services a bit over 8.7 trillion international dollars (i.e., PPP - at 2011 prices). The average person in the world was poor (annual income was somewhat in excess of 3,500 international dollars). Since then the world has prospered beyond recognition. Life expectancy at birth in 1950 was 45, today it is a little over 70. Population has grown to over 7.5 billion and world output of final goods and services is (at 2011 prices) above 110 trillion international dollars, meaning that world income per capita is now more than 15,000 international dollars. A more than 12-fold increase in global output in a 65-year period not only helps to explain the stresses to the Earth system that we have just reviewed, but it also hints at the possibility that humanity's demand for the biosphere's services has for several decades exceeded sustainable levels.⁷

In a review of the state of the biosphere, World Wildlife Fund (2008) reported that although the global demand for ecological services in the 1960s was less than supply, it exceeded supply in the early years of the present century by 50 per cent. The figure is based on the idea of "global ecological footprint," which is the surface area of land and water that would be needed for the Earth system to supply on a sustainable basis the goods and services we consume (food, fibres, wood, water) and to assimilate the waste we produce (materials, gases). The Global Footprint Network (GFN) updates its estimates of the global ecological footprint on a regular basis. A footprint in excess of 1 says demand for ecological services exceeds their supply, which means there is no way for the world to meet our current requirements

⁶ The Anthropocene Working Group has proposed that the immediate post-war years should be regarded as the start of the Anthropocene. See Vosen (2016).

⁷ We are deliberately avoiding the less inexact estimates of current global GDP and population numbers that are available. Below in the text we make use of estimates of global ecological footprints that are even cruder than figures for familiar economic and demographic variables. We want to maintain a balance in our exposition. Undue precision of global figures can mislead.

sustainably. By GFN's reckoning, maintaining the world's average living standard at the present level would require 1.7 Earths.⁸

These are *really* crude estimates, and we feel nervous using them. Figures for such socio-economic indicators as GDP, population size, life expectancy, and adult literacy are reached by a multitude of national and global institutions, who exchange information and coordinate their work. They are rehearsed regularly and governments and international agencies use them routinely when advocating and devising policy. We all take note of their figures and trust them. In contrast, we are obliged here to rely on the estimates of a solitary research group (GFN), albeit one with a network of collaborators. Most people will look askance at their estimates. What matters though is not the exact figure but whether the footprint exceeds 1. On that there should be little question. That there is an overshoot in global demand for the biosphere's goods and services is entirely consistent with a wide range of evidence on the state of the biosphere, which we have reviewed here all too briefly. As the estimates from GFN are the only ones on offer, we make use of them.

Sustainable development would require that the footprint must on average equal 1 over time. Global demand for ecological services can exceed supply for a period, but not indefinitely. Economic development during the past 65 years has raised the average living standard beyond recognition even while population has increased by an unprecedented amount; but we have enjoyed that success by leaving a substantially diminished biosphere for future generations. It would appear we are living at once in the best of times and the worst of times.

It is hard to be sure what a global footprint of 1.7 means for the limits of a sustainable world economy. On the one hand it is worth noting that the Global Footprint Network estimates that the last time we had a footprint close to 1 was in 1970, when the global economy was only 15 % of the size it is today. On the other hand, the size of the footprint has shown limited variation over the past 10 years, despite continued global economic growth, indicating that humanity is starting to operate more efficiently in terms of the amount of consumption that we can generate from the biosphere. Since we will need to estimate a largest sustainable size for the global economy later on in this paper, and believe that it cannot be possible to achieve this via technological innovations on their own, we shall assume a roughly linear longterm relationship between the size of the global economy and its impact on the environment. To err on the cautious side, we work with a rounded figure of 1.5 for the global ecological footprint, which will imply that to reduce the footprint to 1 it will be necessary to reduce the size of the global economy to around 67 % of a 110

⁸ For pioneering work on the idea of ecological footprint, see Rees and Wackernagel (1994) and Rees (2006). See also Kitzes et al. (2008). Wakernagel, who founded the Global Footprint Network (www.footprintnetwork.org/public), was a lead author of WWF (2008).

trillion international-dollar economy. Under the utterly rough approximations we are making here, that would eliminate the current global overshoot in our demands of the biosphere.

4. The Model

Let us then model what these ecological constraints mean for the optimum population. For vividness we think of the stock of global capital assets as the biosphere. Labour, working on the biosphere, can produce output. Our model is of the world as whole. The biosphere is an asset of size K. We may think of K as being measured in units of biomass (tons, say). It could seem as though we are neglecting produced capital (roads, buildings, machines), but to avoid constructing a full-blown economic model we imagine that substitution possibilities between produced capital and the biosphere are extremely limited. So we build the stock of produced capital into K.

Our model is timeless, which means the output is consumed entirely (there is no scope of saving for the future). We investigate the joint optimum population and living standard. The optimization exercise is conducted in two stages. In the first stage the world is populated only by adults, who are assumed to be identical in every way. The latter feature means that any one of them can serve as the decision maker (DM). DM chooses how many children to have, and hence the size of the next generation. These children are identical to the adults in their propensity to consume, but do not have any labour to offer (children don't work!). In the second stage the adult population applies its labour to the biosphere to produce output. That output is consumed equally by the entire population. Even though it is timeless, the model mimics a dynamic model in a stationary mode in which people live for two periods, consuming in both but producing only when they are adults.⁹ Adults in the model are thus both producers and consumers, their offspring are taken to be consumers only.

4.1 Production and Consumption

Let us set the number of people in the global population at N_0 . All these people are identical in every respect and DM is one of them. She acts on behalf of all N_0 people. As we are concerned to study optimum global population, it is natural to assume that all "externalities" have been eliminated: by assumption DM controls the use of the biosphere. As the model is timeless, each individual's life cycle is embedded in it. If *C* is someone's consumption level - we may call it his "living standard" - his personal well-being is U(C). The problem facing DM is to determine the number of people to add. We assume that the world is not overly populated, which means *K* is large rela-

⁹A formal demonstration of that claim is in Dasgupta (2018).

tive to N_0 . It also means that if they were not to have children, each person would enjoy a high standard of living. And if they do have children, they will have to share whatever output they produce among all on an equal basis. So, adding to their numbers will cost the adults, but because they are Generation-Relative Utilitarians, they will still have children if they judge that these children will benefit enough to outweigh these costs.

Let *Q* be output of the consumption good. We first study general production possibilities. If the size of the working population is *N*, we assume that

$$Q = AF(K,N), \qquad A > 0 \tag{1}$$

We want to keep to the simplest assumptions regarding production possibilities. So, we suppose that in equation (1) *F* is homogeneous of degree 1, increasing with *K* and *N*, respectively, at diminishing rates, and F(0,N) = F(K,0) = 0. Figure 1 displays the shape of AF(K,N) as a function of *N*.

A is a parameter in the model, not a variable. In economics it is called "total factor productivity," and it can be interpreted as an aggregate measure of the society's knowledge base and its institutions. Because the model is timeless, we take *K* to be a parameter as well. Notice also that average output per worker (AF(K,N)/N) is a declining function of the number of workers. This is shown in Figure 2.

The *U*-function is assumed to increase with *C* but at a diminishing rate. Write marginal well-being (dU(C)/dC) as U_c , and the marginal of marginal well-being, d^2U/dC^2 as U_{CC} . Thus $U_C > 0$ and $U_{CC} < 0$. Positive well-being (U > 0) records life as good for the person, negative well-being (U < 0) records life as not good. *U* is positive at large values of *C* but negative at small values of *C*. It follows there is a unique value of *C* at which *U* is zero. We write the standard of living at which U = 0 as C^S . Thus $U(C^S) = 0$. Meade (1955) referred to C^S as "welfare subsistence". However, in keeping with our axiology we will refer to it as "well-being subsistence" instead. Earth's carrying capacity is the human population size that the biosphere can support when everyone is at well-being subsistence. Let us label this $2N^S$. When we come to calculate Earth's carrying capacity, it will become evident why we are introducing the factor 2 when labelling Earth's carrying capacity.

The decision facing DM involves two stages. At the first stage she chooses how many children to have, and hence the number of people to be created (N_1). We assume that each of those N_1 people will have the same *U*-function as the N_0 adults. In making that choice DM knows that once born, the additional people will join the existing population as consumers (but not as producers) and enjoy the same status as they. Production and consumption take place in the second stage.

5. Applying Generation Relative Utilitarianism

Equation (1) says that if as the work force is of size N_0 , total output will be $AF(K,N_0)$. As that output will be shared equally among a population of size N_0+N_1 , each person's living standard will be $AF(K,N_0)/(N_0+N_1)$, which is *C*.

Without loss of generality, the weight awarded by DM to the own well-being of the existing population is 1. The weight she awards to the well-being of potential people is μ , where $0 < \mu < 1$. In contrast, Sidgwick's Classical Utilitarianism demands $\mu = 1$.

It will be noticed that Generation-Relative Utilitarianism invokes only a weak form of agent-centred prerogatives. It curbs births, it doesn't sanction applying a lower weight on others' well-being even on grounds of prerogatives. But depending on the relative weights deployed by DM, the gap between ex ante and ex post reasoning can have huge implications for optimum population size.

6. Optimum Population and Consumption

Let *V* be DM's conception of social well-being. Because her perspective plays a role in that conception, we will call *V* her *social valuation function*. Because $C = AF(K,N_0)/(N_0+N_1)$,

$$V = N_0 U(AF(K,N_0)/(N_0+N_1)) + \mu N_1 U(AF(K,N_0))/(N_0+N_1)$$
(2)

6.1 Basics

The problem before DM is to choose N_1 so as to maximize V in equation (2). Routine calculations show that the optimum value of N_1 satisfies

$$U(C) = [(N_0 + \mu N_1) / \mu (N_0 + N_1)] C U_C > 0$$
(3)

The intuition behind equation (3) is this:

At the optimum neither a small hypothetical increase in population nor a small hypothetical decrease would alter DM's social valuation of states of affairs. Suppose now a marginal increase in numbers is contemplated by DM (the argument associated with a marginal decrease in numbers is analogous). The additional person (who is only a potential person in the calculation) would share Q equally with the population that was originally contemplated. The value of that additional well-being as judged by the decision maker would be U(C). But there would also be a decrease in well-being because all others would have slightly less consumption. That potential

loss, per person, is CU_c . Generation-Relative Utilitarianism requires, however, that as there are N₀ actual persons and N₁ potential persons, the effective number of people who would experience that well-being loss per person is $(N_0+\mu N_1)/\mu(N_0+N_1)$. That potential loss in well-being is therefore the expression on the right-hand side of equation (3). At the optimum the potential gain and the potential loss in *V* must be equal. Equation (3) asserts this.

Since Classical Utilitarianism insists $\mu = 1$, equation (3) reduces to the condition

$$U(C)/C = U_C \tag{4}$$

Call the standard of living in equation (4) C^* . Figure 3 shows that at C^* average wellbeing per unit of consumption equals marginal well-being.

In contrast, according to Generation-Relative Utilitarianism $\mu < 1$. In that case $(N_0+\mu N_1) > \mu(N_0+N_1)$. Equation (3) now says that at the optimum, $U(C) > CU_C$. Notice also that the smaller is μ , the greater is the difference between U(C) and CU_C , which in turn implies that the optimum standard of living is higher. This is demonstrated in Figure 3.

6.2 Reproductive Replacement

We now consider an extreme scenario. Imagine that the optimum policy for the existing population is to replicate itself. That's when N_0 just *happens* to be a figure for which optimum N_1 equals N_0 . The reason we are interested in this extreme scenario is that it mimics an economy moving through time in a *stationary* mode. The stationary state in question would be one where people live for two periods and produce children at the beginning of their second period. Being stationary, adults in the economy are replaced by an equal number of children in each period.

So, in our timeless economy, we set $N_0 = N_1 = N$. Total population is then 2*N*, and the optimality condition (eq. (3)) becomes

$$U(C)/C = [(1+\mu)/2\mu]U_C > 0$$
(5)

Let the solution of equation (5) be denoted as C^{0} . That is to say C^{0} is the optimum living standard under Generation-Relative Utilitarianism.

Since Classical Utilitarianism demands that μ = 1, equation (5) again reduces to the condition

$$U(C)/C = U_C \tag{6}$$

We have dubbed the solution of equation (6) as C^* . A comparison of equations (5) and (6) tells us immediately that $C^0 > C^*$.

It will pay to study the optimum population-consumption mix by using simple forms of the production function AF(K,N) and the well-being function U(C). Their simplest expressions are power functions. So, we work with them to study equation (5). Let

$$AF(K,N) = AK^{(1-\rho)}N^{\rho}, A > 0, 0 \le \rho < 1$$
(7)
$$U(C) = B - C^{-\sigma}, B > 0, \sigma > 0 (8)$$

Equation (7) is widely used by economists to reflect production possibilities. The parameter ρ reflects the productivity of labour. Output is an unbounded function of population numbers, but Nature imposes a restraint on the rate at which output can expand with population. The latter is reflected in the condition $\rho < 1$. (1- ρ) is the productivity of *K*.

Ideal national income accounting would interpret ρ to be the share of total output attributable to labour. There is an enormous empirical literature offering estimates of ρ . They tend to lie in the range 0.6-0.7. Because we are including the biosphere in the accounts, ρ should be taken to be smaller. In numerical exercises we will assume, solely for computational ease, that $\rho = 0.5$.

U(C) in equation (8) is defined by two parameters, *B* and σ , both positive numbers. *U* is bounded above, the least upper bound being *B*. Ramsey (1928) called *B* Bliss. $1+\sigma$ is the absolute value of the percentage rate at which marginal well-being changes with each percentage rate of increase in consumption (i.e. $1+\sigma = -d\log(U_C)/d\log C$). Which is why $1+\sigma$ is called the "elasticity of marginal well-being with respective to consumption." The elasticity exceeds 1 in the *U*-function of equation (8). It is immediate if U(C) satisfies equation (7), then $C^S = B^{-1/\sigma}$.

Sidgwick's Utilitarianism requires μ = 1. In that case equations (6) and (8) yield

$$C^{*}/C^{S} = [1+\sigma]^{1/\sigma} < e = 2.7...$$
(9)

Equation (9) tells us that the optimum standard of living isn't much greater proportionately than well-being subsistence (Dasgupta, 1969).

In contrast, there is no upper bound to the optimum living standard under Generation-Relative Utilitarianism. Intuition suggests that the smaller is μ , the larger is C° . That is indeed so. To confirm, recall that we defined Earth's carrying capacity as

the size of the global population that the biosphere could sustain at living standard equal to well-being subsistence. We denoted the size of that population by $2N^{S}$. Thus,

$$AK^{(1-\rho)}(N^S)^{\rho}/2N^S = C^S$$
⁽¹⁰⁾

Using equations (7)-(8) in equation (5) and using equation (10) yields

$$C^{O}/C^{S} = (N^{S}/N^{O})^{(1-\rho)} = [1+(1+\mu)\sigma/2\mu]^{1/\sigma}$$
(11)

As expected, comparison of equations (9) and (11) shows that $C^{O} > C^{*}$. This is shown in Figure 3. Moreover, equation (11) says that the smaller is μ , the larger is C^{O}/C^{S} .

In order to obtain a quantitative sense of the optimum under Generation-Relative Utilitarianism, we need numerical values of ρ , σ and μ . We assume $\rho = 0.5$. σ has been estimated from consumption behaviour under risk. $\sigma = 1$ is at the upper end of the range that has been found in empirical studies. For ease of computation we settle on that.

Stopping rules that are used by households to determine their family size could in principle be used to estimate u, but we don't know of any study that has gone that route. No doubt household behaviour isn't the exclusive source of ethical understanding, but it would be wrong to ignore people's intentions altogether in reaching ethical directives. Casual empiricism on health and education expenditures in time and money on children in the West, especially perhaps on children with special needs, suggest that μ is considerably less than 1. People seem to place far greater weight on the well-being of their children than on the potential well-being of children who might have been born but weren't because couples chose not to have further children. For illustration, let us assume $\mu = 0.05$. This is to take a lunge in the dark, but we are using the figure only for illustrative purposes. Using the figures in equation (9) yields $2N^S/2N^O \approx 132$, meaning that the biosphere's carrying capacity for humans is (approximately) 132 times optimum population. This is a far cry from the ratio for Total Utilitarianism, which is close to 1:1. Suppose C^{S} is taken to be below even the average annual income in the world's low-income countries; say 1,500 international dollars (see Table 1). In that case equation (9) says $C^0 = 17,250$ international dollars. That's somewhat above the global per capita income today of 15,000 international dollars and would be regarded by the World Bank as a Middle Middle-Income Country.

7. Earth's Human Carrying Capacity

To determine optimum population $2N^{O}$, we need an estimate of $2N^{S}$, and to obtain the latter, we need data on the biosphere's productivity and its regenerative possibilities. There is little quantitative information on the biosphere's dynamics, but we know from the historical experience that expanding our stock of produced capital is likely to have adverse environmental consequences. So, as stated previously, with both hands proverbially tied behind our backs we regard K to be an aggregate measure of the biosphere and produced capital. Next, we stop both A and K on their tracks and estimate $AK^{(1-\rho)}$ (eq. (7)) on the basis of figures for the global ecological footprint, the current size of the world economy, and our model of global production (eq. (7)). The estimate of $AK^{(1-\rho)}$ includes the social value not only of the biosphere, but also of produced capital, social institutions, and public knowledge. In short, we use equation (6) to estimate the social worth of all capital assets with the exception of labour. That enables us to estimate $2N^{S}$. We appreciate that is applying an intellectual sledge-hammer to a delicate problem (even if humanity were to disappear from the face of the earth, the biosphere's dynamics would be shaped by the human imprint of the past), but we have found no other way to get at $2N^{S}$.

The data being utterly crude, we confine ourselves to pen-on-paper computations. We assume that the value of the world's production of final good and services draws proportionately on ecosystem services at all levels. World output is currently about 116 trillion international dollars. Let 2*N* denote the global population. So as to remain in step with our timeless model, we assume that half of our numbers are engaged in production. Using equation (7) yields

$$AK^{1-\rho}(N)^{\rho} = 116 \text{ trillion (international) dollars}$$
 (12)

We take world population to be 7.6 billion, and we continue to assume $\rho = 0.5$. Write $AK^{1-\rho}$ as **K**. Equation (12) then says

$$\boldsymbol{K} = 116 \times 10^{12} / (3.8 \times 10^9)^{0.5} \text{ dollars per producer}^{0.5}$$

$$\approx 1.9 \text{ billion dollars per producer}^{0.5}$$
(13)

We assume there is a roughly linear relationship between the size of our economy and its impact on the world. We also assume that the global economy currently has a footprint of 1.5 earths. That means if the biosphere and all other forms of capital assets barring labour were to be stopped on their tracks, their sustainable value would be K/1.5, which we denote by K^* . Using equation (13),

$$K^* \approx 1.2$$
 billion dollars per producer^{0.5} (14)

We have now calibrated the model and have all the information we need to estimate $2N^{S}$ and therefore $2N^{O}$.

In our model,

$$\boldsymbol{K}^{*}(N^{S})^{\circ}/2N^{S} = C^{S}$$

$$\tag{15}$$

Suppose, as in the previous numerical example, DM sets $\rho = 0.5$ and $C^S = 1,500$ international dollars. Then applying the two numerical figures in equations (14)-(15) yields $2N^S = 320$ billion. That's Earth's carrying capacity. As before, we assume DM sets $\mu = 0.05$ Equation (11) then yields $2N^O = 2.4$ billion. That's the optimum population size. Classical Utilitarianism in contrast would recommend a global population of 80 billion! Once again, this confirms Classical Utilitarianism's preference for larger population with lower standards of living than Generation-relative Utilitarianism' recommendations.

Part III. Reconciling Generation Relative Utilitarianism to Global Inequality and Differential Fertility

8. Decentralizing the Optimum

In formulating the problem of optimum population we assumed in Part II that the DM is able to enforce the *full optimum*, by which we mean that the only constraints she faces are those on aggregate production possibilities (eq. (1)). Because people having identical *U*-functions, the DM quite rightly does not discriminate among households in any way. To achieve a full pptimum, the DM would equalise household wealths by imposing lump sum wealth transfers. Economists call the full optimum the "first best."

It will have been noticed from equation (11) both C^O/C^S and N^O/N^S are independent of K^* . This is a striking feature of the Utilitarian calculus, be it the Classical or the Generation-Relative version. Well-being subsistence is an expression of a pure value judgment and does not depend on production possibilities. K^* is in contrast an expression of the economy's production possibilities. So, equation (11) says that the livings standard at the full optimum is independent of K^* . On the other hand, equation (10) says that Earth's carrying capacity, $2N^S$, depends both on C^S and

the economy's production possibilities. For any chosen value of C^S , the larger is K^* , the larger is Earth's carrying capacity. Correspondingly, the optimum population is larger by the same proportion.

If we reinterpret the model in terms of households, our analysis says that if all households choose the number of children to have using the same value of well-being subsistence C^{S} , then wealthier households will have greater numbers of children. Unfortunately, this does not cohere with the observations of economic demography, since it is clear that not only do different people choose to have different numbers of children, but that there is a persistent trend for poorer families, on average, to have more children than richer families. Table 1 illustrates this phenomenon. Of course, this trend hides many important counterexamples, for instance amongst some rapidly developing poor countries with low rates of fertility and some very wealthy families who do appear to use some of their wealth to have more children than their relatively poorer peers. However the general trend remains a problem for our model because whilst it is easy to say that people are, for a variety of reasons, simply making the wrong choices, if peoples choices are actually inverse to what our model would suggest they should be then either this is a sign that we have failed to capture some ethically salient feature of these people's situation, or else that no amount of ethical reasoning is going to be able to effect significant change on them.

Among economists a now-traditional explanation for the observation that wealthier households have fewer children is based on the (market) value of time (Becker, 1960; Becker, Murphy, and Tamura, 1990). Wealthier households are wealthier because, or so the argument goes, their wages are higher. And their wages are higher because they have acquired more human capital (e.g., women's education) than less wealthy households. So, the value of time is higher for wealthier households. If you now acknowledge that bearing and rearing children, taken together, is very time consuming, you have an explanation. But the explanation is restricted to market economies. The Beckerian framework does not work well in poor countries, because rural women there are required to do a huge amount of work each day. In this section and the next we will therefore consider two possible factors that may help explain this apparent discrepancy between what our model says about the optimum behaviour for DMs and what people are actually choosing to do. For a review of the many other social factors that affect fertility decisions in poor countries, see Dasgupta (1993).

8.1 The Biosphere as a Differentiated Commodity

Adverse externalities arising from our use of the biosphere in great measure arise because Nature is mobile: birds and insects fly, water flows, the wind blows, and the oceans circulate. That makes it hard to establish property rights to key components of the biosphere. By property rights we don't only mean private rights, we include communitarian and public rights. Which is why much of the biosphere is an "openaccess resource," meaning that it is free to all to do as they like with it. Hardin (1968) famously spoke of the fate of non-managed common property resources as "the tragedy of the commons." But while Hardin's analysis was entirely appropriate for global commons (the atmosphere, the oceans), it was less than applicable to geographically confined resources such as woodlands, ponds, grazing fields, coastal fisheries, wetlands, and mangroves. Because local commons are geographically confined, their use can be monitored by community members. There were exceptions of course, but in times past those resources were managed by communities, they were not open-access resources. Reviewing an extensive literature, Feeny et al. (1990) observed that community management systems enabled societies to avoid experiencing the tragedy of the commons. Social norms of behaviour, including the use of fines and social sanctions for misbehaviour have guided the use of local common property resources.¹⁰

In poor countries the commons continue to supply household needs and marketable resources to rural people (water, fuelwood, medicinal herbs, fruits and berries, manure, wood and fibres and timber for building material). As in so many other spheres of social life, communitarian practices have over the years strengthened in some instances (e.g. community forestry in Nepal) and weakened in others, especially when communal rights were overturned by central fiat. An example of that latter was when, in order to establish political authority after independence (and also earn rents from timber exports), a number of states in sub-Saharan Africa and Asia imposed rules that destroyed community practices in forestry. However, knowledge of local ecology is still largely held by those who work on the commons, not by state officials, who in addition can be corrupt. Thomson et al. (1986), Somanathan (1991), and Baland and Platteau (1996), among others, have identified ways in which state authority damaged local institutions and turned local commons into seemingly open-access resources. Then there are subtle ways in which even wellintentioned state policy can cause communitarian practices to weaken (Balasubramanian, 2008; Mukhopadhyay, 2008).11

¹⁰ The literature on this is extensive. See Ostrom (1990), Marothia (2002), Ostrom et al. (2002), and Ostrom and Ahn (2003).

¹¹ In recent years democratic movements among stakeholders and pressure from international organisations have encouraged a return to community-based systems of management of the local commons. Shyamsundar (2008) is a synthesis of the findings in nearly 200 articles on the efficacy of a devolution of management responsibilities - from the state to local communities - over the local natural-resource base. Her article focuses on wildlife, forestry, and irrigation. The balance of evidence appears to be that devolution leads to better resource management, other things equal. Shyamsundar of course offers a discussion of what those other things are.

There is little quantitative evidence to show how important local commons are to household income. Casual empiricism suggests they are less significant in advanced industrial countries than in poor rural societies. In the former, local resources are either owned privately, fall under the jurisdiction of local authorities or, as in the case of places of especial aesthetic value, are national parks. That is not so in rural areas in poor countries. In a pioneering work, Jodha (1986) reported evidence from semi-arid rural districts in Central India that among poor families the proportion of income based directly on local commons was 15–25 %. Cavendish (2000) arrived at even higher estimates from a study of villages in Zimbabwe: the proportion of income based directly on local common property resources was found to be 35 %, the figure for the poorest quintile being 40 %. To not recognize the significance of the local natural-resource base in poor countries is to not understand how the poor live.

Even when a common is managed by the community and outsiders are kept at bay access to it may be based on household size or each household may have a fixed share of its output. When larger households are entitled to a greater share of the commons' goods and services, households have an incentive to convert natural resources excessively into private assets, including size. In sub-Saharan Africa larger households are (or until recently, *were*) awarded a greater quantity of land by the kinship group. What is true in the case of local commons to which households have access regardless of their size holds true in the case of global commons, to which we all have access regardless of our household size. Why might therefore expect DMs choices about family size to reflect both the degree to which they depend upon commons for their households' access to resources and the ways in which these commons are managed.

Production possibilities in equation (7) possess the property of "self-similarity," which is a way of saying that every feasible mix of inputs in production can be applied at any scale of operation. Assuming that access to *K* can be monitored and restricted, we first uncover (Sect. (8.2)) the system of property rights to *K* that can implement the full optimum (eq. (11)) in a decentralized world. In Section (8.3) we study fertility choice when a household's entitlement to *K* increases with household size. We find that this form of property rights motivates households to reproduce more than at the full optimum.

8.2 Complete Property Rights to the Biosphere

As before, the model is timeless. There are N_0 adults who are the producers and decision makers. By mutual agreement, households are taken to be the social units irrespective of their size. So, each individual is restricted to the fraction $1/N_0$ of the common property resource *K*. We study the decisions of person *i* (*i* = 1,..., N_0).

Let *n* be the number of children each of the remaining people intend to have, and let n_i be the number of children *i* intends to have. If intentions are realized, total population will be $[(N_0-1)(1+n)+(1+n_i)]$. In view of the entitlement rule we are studying, the quantity of *K* to which each household has access is K/N_0 . Without loss of generality we set A = 1 in equation (7). If C_i denotes average consumption in *i*'s household,

 $C_i = (K/N_0)^{(1-\rho)}(1+n_i)^{-1}$

That means person i's objective is to maximize V_i , where

$$V_i = U((K/N_0)^{(1-\rho)}(1+n_i)^{-1}) + \mu n_i U((K/N_0)^{(1-\rho)}(1+n_i)^{-1})$$

or

$$V_i = (1 + \mu n_i) U((K/N_0)^{(1-\rho)} (1 + n_i)^{-1})$$
(16)

It follows from equation (16) that i's optimal choice of n_i satisfies the condition

$$\mu U(C_i) = [(1 + \mu n_i)(1 + n_i)^{-1}]C_i dU(C_i)/dC_i$$
(17)

In a *social equilibrium i*'s optimal choice of n_i equals n. Moreover, in the replicating economy N_0 just happens to be the number for which n = 1. Denote that number as N^{OO} . Population size is then $2N^{OO}$. Let C^{OO} be average consumption per person in social equilibrium. Setting $n_i = n = 1$ in equation (17) implies that C^{OO} satisfies

$$U(C) = [(1+\mu)/2\mu]CU_C$$
(18)

But equation (18) is the same as equation (5). Similarly, we confirm that fertility behaviour in social equilibrium satisfies equation (5), which means it is fully optimal. That proves that by parcelling the biosphere equally among households the decentralized economy achieves the fully optimum living standard and population size. That is, $C^{O} = C^{OO}$ and $N^{O} = N^{OO}$.

8.3 Open-Access Resources and Population Overshoot

We now consider a different property-rights regime. Imagine that the community takes household size into account when allocating rights to the commons. In that case *i*'s household would be entitled to $\{(1+n_i)/[(N_0-1)(1+n)+(1+n_i)]\}$ -th portion of *K*. To avoid clutter, write $H = (N_0-1)(1+n)$. Then in place of equation (16), we have

$$V_i = (1 + \mu n_i) U((K/(H + 1 + n_i))^{(1 - \rho)})$$
(19)

In a social equilibrium *i*'s optimal choice of n_i equals *n*. And as in the replicating population studied previously, we assume N_0 just happens to be the number for which n = 1. Denote that number as N^{**} . Population size is then $2N^{**} > 1$. Let C^{**} denote average consumption per person in social equilibrium. From equation (19) it follows that C^{**} satisfies

$$U(C) = [(1+\mu)(1-\rho)/2\mu] C U_C/2N^{**}$$
(20)

For vividness, suppose U is the power function in equation (8). Then equation (20) reduces to

$$C^{**}/C^{S} = (N^{S}/N^{**})^{(1-\rho)} = \{ [1+\sigma(1+\mu)(1-\rho)/2\mu]^{1/\sigma} \}/2N^{**}$$
(21)

Comparison of equations (11) and (21) shows that $C^{**} < C^{OO}$ and $N^{**} > N^{OO}$. The result confirms that when increasing family size becomes a means of increasing ones share of the commons then this leads to a higher optimal rate of fertility: Freedom on the commons leads to over-population.

9. Working with Non-Uniform Conceptions of Value

When comparing Generation Relative Utilitarianism and Classical Utilitarianism, it was noted that the optimum population size and its associated standard of living depend significantly on one's agent centred prerogatives. In particular we saw that the value of μ plays a significant role in determining this. If, as classical utilitarianism demands, we set μ as equal to 1 then optimum population size would be 80 billion. However, if we set it as equal to, say, 0.05, then optimum population falls to 2.4 billion.

It appears to us that the correct value for μ is not yet amenable to direct ethical analysis, but rather that it reflects current cultural beliefs about the importance of

individual persons and our attitude to time. We have argued elsewhere for the moral correctness of a relatively low value for μ (Dasgupta 2018, Beard forthcoming). However, it is likely that such a view is more reflective of contemporary, individualistic, norms that have come to dominate many societies since early modern Europe. Others have argued that in traditional societies people's values are often more communitarian and timeless, suggesting that the value of μ that would reflect observed behaviour in such societies could be considerably closer to one. (Metz 2007)

Any population in which households possess different values for μ can display a variety of behaviours. If property rights are well established then these may stratify into different groups, each at the optimum standard of living for them and their children implied by their own value of μ and a population size that reflects this value along with that groups share of global resources. In a more open society in which households gain at least some additional access to resources based on their family size however, households with lower values of μ can be expected to gain a larger and larger share of these resources, so that global population will tend towards that associated with the highest value for μ of any household.

Contemporary global societies undoubtedly display at least some degree of openness. However, the picture is complicated by the fact that household values for μ appear to change over time depending on a wide range of factors. These include cultural and religious norms and economic circumstances. However, they also reflect underlying environmental conditions, with uncertain situations like wars and pandemics showing a demonstrable effect of raising fertility rates, potentially because they encourage people to take a more communitarian and timeless approach to value in general, or merely to give less weight to the interests of presently existing people (who may soon die) over potential future people (who may be more likely to survive this present crisis; see e.g., Pepper et al. 2016).

A second feature of our modal that is likely to vary across different societies are the levels that we have referred to as well-being subsistence. Recall that this refers to the level of consumption at which a person's life will be neither good nor bad. This factor can vary according to a variety of different factors including the relative purchasing power of money in these societies, the different costs of 'doing business' in these communities (such as meeting the basic standards of acceptability for being granted full membership of its social and economic institutions. For reasons that have been much studied by anthropologists, people's preferences are socially embedded. In traditional societies there is greater conformity. Taken together with the factors we studied in Section 8.3, we may infer that households would be larger in poorer societies.
10. Resolving Disagreements about the Optimum Population

These are important findings, because they help to explain an apparent discrepancy between the normative implications for what decision makers should choose to do in terms of moving towards an optimum population, and real-world observations about what moral agents actually do in the real world. However, this distinction should not be taken too far. For one thing it has been widely observed that many people in developing countries still do not have effective control over their reproductive decision making, due to a lack of education, a lack of access to reproductive health or a fatalistic mentality that is dismissive of the possibility for human agency over reproductive decisions. When all these barriers to individual decision making have been overcome this often goes a long way towards bringing people's attitudes towards reproductive decision making into line, even when they come from differrent sociocultural communities.

However, our model would seem to imply that even once these more practical issues have been dealt with there remains scope for a genuine disagreement about optimal reproductive behaviour based on social and cultural practices. Depending upon their attitudes to time, their view about the value of as yet unborn people's welfare and their sense of the minimum standard for a life's being worth living, reproductive decision makers will, it seems, have reason to substantially disagree about the world's optimum population, and hence their optimum family size and reproductive decision making. Moreover, aspects of this disagreement do appear to boil down to a genuine ethical difference between people that may not be amenable to any simple or reductionist resolution into facts about the world from a demographic or environmental perspective.

Nevertheless, we conclude that certain facts now facing humanity should rightly be seen as having a bearing on this debate, and that they provide the grounding for a robust defence of lower assessments for the global optimum population. The first of these relates to the probability that future generations will continue to be born at all, or at least that they will continue to be born in anything like the numbers, and with anything like the life prospects, of the present. Unfortunately, the global population is not simply a contributor to our overall consumption of the earth's resources, but is also a direct contributor to our level of existential threat (Ehrlich and Ehrlich 2013, Kuhlemann 2018). Whilst it is possible to support larger and smaller global populations, larger populations will invariably require more infrastructure to support them and this will have a direct impact on the resilience of this infrastructure to external shocks, both environmental and social. Larger populations also make the task of adapting to large scale global changes, whether technological or environmental, more challenging since more work must be done instantaneously without the opportunities for learning. Previous population growth means that somewhere in the region of 15 % of humans who have ever lived are currently alive, whilst the combination of this growth with technological change means that around 90 % of all scientific research has been undertaken by living scientists. This naturally limits the opportunity for humans to coordinate their actions and to learn from one another and it would be better for our species if in the future the population where to stabilize or shrink so that we could make wiser decisions for the benefit of our species.

A second crucial consideration is the natural striving for every living person to better themselves if they can. Whilst parents may chose to have more children even though this may make these children less well off, we cannot expect children to bind their expectations in life to the judgement of their parents. Children and adults everywhere can legitimately be expected to want to have as good a life as possible for themselves. This will make overconsumption of the earth's resources far harder to avoid under situations in which the global population is larger, and people are expected to stop consuming at a lower quality of life, than if it is lower and they are allowed to consumer more without exceeding the planets carrying capacity. Furthermore, when resources are scarce and people's welfare is being held far below what it could be, and especially under conditions of inequality, this is likely to foster conflict.

To the extent that avoiding human extinction and preventing conflict are universally desirable we therefor conclude that disagreements about optimum population and reproductive decision making should not simply be left unresolved, but that there are reasons to promote engagement with the aim of encouraging smaller family sizes and more equal reproductive behaviour around the world. However to the extent that our model is able to explain differential behaviour it also offers hope for how to respond to this effectively and without coercion, by engaging with the substantial moral beliefs that may in part be driving larger family sizes. It does not strike us as improper for scholars to note that if having smaller families is not only of benefit to children but also benefits all of humanity then it is a good thing in itself, and to be willing to stand up for this view even though it may at times be controversial. So long as we stick to the facts about why global population and reproductive decision-making matters in the world today then we should have nothing to fear.

References

Baland, J.-M. and J.-P. Platteau (1996), *Halting Degradation of Natural Resources: Is There a Role for Rural Communities*? (Oxford: Clarendon Press).

Balasubramanian, R. (2008), "Community Tanks vs Private Wells: Coping Strategies and Sustainability Issues in South India", in R. Ghate, N.S. Jodha, and P. Mukhopadhyay (2008), *Promise, Trust and Evolution: Managing the Commons of South Asia* (Oxford: Oxford University Press).

Beard, S. (Forthcoming) *Fairness Based Arguments for a Positive Pure Time Preference*, ms.

Becker, G.S. (1960), "An Economic Analysis of Fertility", in *Demographic and Economic Change in Developed Countries*, NBER Publication (New York: Columbia University Press).

Becker, G.S., K.M. Murphy, and R. Tamura (1990), "Human Capital, Fertility, and Economic Growth", *Journal of Political Economy*, 98(5), S12–S37.

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

Cavendish, W. (2000), "Empirical Regularities in the Poverty-Environment Relationships of Rural Households: Evidence from Zimbabwe", *World Development*, 28(11), 1979–2003.

Dasgupta, A. and P. Dasgupta (2017), "Socially Embedded Preferences, Environmental Externalities, and Reproductive Rights", *Population and Development Review*, 43(3), 405–441.

Dasgupta, P. (1969), "On the Concept of Optimum Population", *Review of Economic Studies*, 36(3), 295–318.

Dasgupta, P. (2018), "Birth and Death - Arrow Lecture", in *Time and the Generations: population ethics for a diminishing planet* (New York: Columbia University Press), forthcoming 2019.

Ehrlich, P. R., & Ehrlich, A. H. (2013). Can a collapse of global civilization be avoided?. Proceedings of the Royal Society B: Biological Sciences, 280(1754), 20122845.

Feeny, D., F. Berkes, B.J. McCay, and J.M. Acheson (1990), "The Tragedy of the Commons: Twenty-two Years Later", *Human Ecology*, 18(1), 1–19.

Griffin, J. (1986), *Well-Being: Its meaning, measurement, and moral importance* (Oxford: Oxford University Press).

Hardin, G. (1968), "The Tragedy of the Commons", Science, 162, 1243-1248.

Jodha, N.S. (1986), "Common Property Resources and the Rural Poor", *Economic and Political Weekly*, 21, 1169–81.

Kitzes, J., M. Wakernagel, J. Loh, A Peller, S. Goldfinger, D. Cheung, and K. Tea (2008), "Shrink and Share: Humanity's Present and Future Ecological Footprint", Philosophical Transactions of the Royal Society, B, 363(25 July), 467–475.

Kuhlemann, Karin. "Complexity, creeping normalcy and conceit: sexy and unsexy catastrophic risks." *Foresight 21*, no. 1 (2019): 35–52.

Marothia, D.K., ed. (2002), *Institutionalizing Common Pool Resources* (New Delhi: Concept Publishing Company).

Meade, J.E. (1955), Trade and Welfare (Oxford: Oxford University Press).

Metz, T. (2007), "Toward an African Moral Theory", Journal of Political Philosophy, 15(3), 321–341.

Millennium Ecosystem Assessment (2008), *Ecosystems and Human Well-Being*, Vols. 1-V (Washington, DC: Island Press).

Mukhopadhyay, P. (2008), "Heterogeneity, Commons, and Privatization: Agrarian Institutional Change in Goa", in R. Ghate, N.S. Jodha, and P. Mukhopadhyay, eds., (2008), *Promise, Trust and Evolution: Managing the Commons of South Asia* (Oxford: Oxford University Press).

Ostrom, E. (1990), *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge: Cambridge University Press).

Ostrom, E. and T.K. Ahn, eds. (2003), *Foundations of Social Capital* (Cheltenham: Edward Elgar).

Ostrom, E., T. Dietz, N. Dolsak, P.C. Stern, S. Stonich, and E.U. Weber, eds., (2002), *The Drama of the Commons* (Washington, DC: National Academy Press).

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

Pepper, G., McAllister, L., & Sear, R. (2016). "Why demography needs psychologists". *The Psychologist*, 29(1), 26–29.

Ramsey, F.P. (1928), "A Mathematical Theory of Saving", *Economic Journal*, 38(4), 543–559.

Rawls, J. (1972), A Theory of Justice (Oxford: Oxford University Press).

Rees, W.E. (2006), "Ecological Footprints and Biocapacity: Essential Elements in Sustainability Assessment", in J. Dewulf and H.V. Langenhove, eds., *Renewable-Based Technology: Sustainability Assessment* (Chichester, UK: John Wiley & Sons).

Rees, W.A. and M. Wackernagel (1994), "Ecological Footprints and Appropriated Carrying Capacity: Measuring the Natural Capital Requirements of the Human Economy," in A.M. Jansson et al., *Investing in Natural Capital: The Ecological Economics Appropriate for Sustainability* (Washington, DC: Island Press).

Scheffler, S. (1982), The Rejection of Consequentialism (Oxford: Clarendon Press).

Shyamsundar, P. (2008), "Decentralization, Devolution, and Collective Action," in R. Ghate, N.S. Jodha, and P. Mukhopadhyay, eds. (2008), *Promise, Trust and Evolution: Managing the Commons of South Asia* (Oxford: Oxford University Press).

Sidgwick, H. (1907), The Methods of Ethics (London: MacMillan), 7th Edition.

Sodhi, N.S., B.W. Brook, and C.J.A. Bradshaw et al. (2009), "Causes and Consequences of Species Extinctions," in S.A. Levin et al., eds., *The Princeton Guide to Ecology* (Princeton NJ: Princeton University Press).

Somanathan, E. (1991), "Deforestation, Property Rights and Incentives in Central Himalaya," *Economic and Political Weekly*, 26 (Special Issue: January 26) PE37–46.

Thomson, J.T., D.H. Feeny, and R.J. Oakerson (1986), "Institutional Dynamics: the Evolution and Dissolution of Common Property Resource Management," in *National Research Council, Proceedings of Conference on Common Property Resource Management* (Washington, DC: US National Academy of Sciences).

Vosen, P. (2016), "Anthropocene Pinned Down to Post War Period," *Science*, 353(6302), 852–853.

Waters, C.N., J. Zalasiewicz, C. Summerhayes, A.D. Barnosky, C. Poirier, A. Galuszka, A. Cerreta, M. Edgeworth, E.C. Ellis, C. Jeandel, R. Leinfelder, J.R. McNeill, W. Steffen, J. Syritski, D. Vidas, M. Wagreich, M. Williams, A. Zhisheng, J. Grineveld, E. Odada, N. Oreskes, and A.P. Wolfe (2016), "The Anthropocene is Functionally and Stratigraphically Distinct from the Holocene," *Science*, 351(6269), aad2622-(110).

World Wildlife Fund (WWF) - (2008), *Living Planet Report 2008* (Gland:WWF International) World Wildlife Fund (WWF) – in collaboration with the London Zoological Society (2018), *Living Planet Report 2018* (Gland: WWF International).

M.A. Roberts¹

Anonymity and Indefinitely Iterated Addition and Reversal²

This paper argues that an unrestricted form of the principle of anonymity in combination with what seems an innocuous addition of a worth-having existence sometimes generates results that are surely false. A more plausible account of the problem case, according to this paper, inductively extends the idea that the morally worse future must be worse for a person who does or will exist in that future (the "existence condition") on grounds of a result previously established by appeal to an adaptation of a certain standard Pareto principle. Along the way, the paper rejects the unrestricted form of anonymity.

¹ The College of New Jersey, robertsm@tcnj.edu.

² An earlier version of this paper was presented at a workshop organized by Julia Mosquera and Anders Herlitz for the Institute for Futures Studies on Claims and Distributive Justice, Stockholm, June 2021. The current version of this paper owes much to the extremely valuable comments made by the participants of that workshop. A more detailed inquiry into the anonymity puzzle is included in Roberts (manuscript). Financial support from Riksbankens Jubileumsfond (grant M17-0372:1) is gratefully acknowledged.

1. Simple and indefinitely iterated addition and reversal

1.1 When addition and reversal seem not to matter

As Parfit's work attests, when principles that work nicely on their own are forced to work together in the odd case—that odd case itself often being an *additional person* case—some very deep puzzles can arise.

Here, we'll start with a case that combines (i) an additional worth-having existence with (ii) what we can call a *merely reversing* change, where both the addition and the reversal may well strike us as completely innocuous.

We'll call it simple addition and reversal (figure 1.1).³

choice	c1	c2	c3	
probability of	1	1	1	
effects as				
described,				
given choice				
wellbeing	f1	f2	f3	
+10	Polly	Polly	Quentin	
+9		Quentin	Polly	
+0	Quentin			

Figure 1.1 Simple addition and reversal

In this and later figures, a name in bold means that the person named does or will exist, and a name in italics means that the person named never exists, in the indicated possible future. The choices (c1, c2, c3) available to the agents give rise, respectively, to distinct possible futures (f1, f2, f3), with each future being accessible relative to each other future.⁴ Numbers in the far left column show overall (lifetime) wellbeing

³ What I here call the simple addition and reversal case was outlined by Mark Budolfson in a workshop organized by Dean Spears on Risk and Population, Department of Economics, University of Texas at Austin, November 2019. The case that generates the problem, or puzzle, I shall address in this paper—the case of the indefinitely iterated addition and reversal, outlined below—is my own. I am grateful to Budolfson and Spears for the workshop itself and ensuing discussion which led me to develop my own solution to what I shall call the *anonymity puzzle*. That solution is set forth toward the end of this paper (part 3 below).

⁴ To say that one future is accessible relative to another is just to say that the agents in the one future have the ability, the power, the resources (as of the time just prior to choice) to bring about the effects described for each other future. A future that agents collectively, working together, have the ability, the power, the resources to bring about counts as accessible for purposes here. That's so, even if, in point of

levels for each person, with wellbeing itself being left undefined for purposes here but understood as a measure of just how precious the particular existence happens to be for the person who has it.⁵

Now, as far as I can see, this case doesn't immediately give rise to any earth-shattering puzzle at all. For it seems—at least at first glance—clear what we should say about this case—and it seems clear—again, at first glance—that what we want to say about this case isn't problematic.

To start, if we think, as a matter of clear intuition, that the addition of a given person to a given future isn't the sort of thing that makes things morally better—if we think, that is, that we make things morally better not by "making happy people" but rather by "making people happy"—then we will want to say that f2 isn't morally better than f1, and nor is f3.⁶

A principle we can call the *existence condition* (EC) supports that intuition.

Existence condition (EC): Where x and y are possible futures and y is accessible relative to x,

x is morally worse than y, and a choice c made at x is wrong, *only if* there is a person p and an alternate accessible future z such that:

p does or will exist in x and

x is worse for p than z (where z may, but need not, be identical to y).⁷

fact, agents fail to cooperate in any way at all and even if the probability that a given future will unfold under the agents' collective choice is very low.

⁵ Wellbeing should thus be understood as raw wellbeing, unadjusted (for purposes of any aggregative function) by, for example, the values of equality, fairness or priority—or by what we can call our existential values.

The term person is also left undefined for purposes here, but should be understood to include many non-human animals and to exclude, at the same time, many humans, with consciousness itself as an indicator of personhood.

⁶ Narveson 1976, p. 73.

⁷ The existence condition (EC) articulates a very narrow version of the person-affecting intuition, itself made famous as a target of Parfit's nonidentity problem. See Parfit 1987, pp. 351-379 and especially p. 363 ("what is bad must be bad for someone"). I abandon any use of the terms "person-affecting" and "person-based" here, however, due to the fact that they have been used to reference a number of principles that are themselves clearly false. In contrast, EC, being very narrowly drawn, provides only a necessary and not a sufficient condition on worseness and provides a condition only on worseness and not betterness. Moreover—and this is a feature helpful in addressing, among others, the nonidentity problem itself—EC's necessary condition on when one future x is worse than an alternate accessible future y is notably satisfied (meaning that EC itself remains silent) in any case in which there exists an alternate accessible future z that is better for a person p who does or will exist in x than x itself is, where z itself may, but need not, be identical with y.

According to EC, the fact that Quentin *never exists* in fl doesn't make fl *worse* than f2 or f3. fl is at least as good as f2 and as f3; and it's perfectly permissible for agents to choose to leave Quentin out of existence altogether.

But what about the other way around? Is f2—or f3—worse than f1? As to those two questions, EC remains silent. It's silent on whether f2 is worse than f1 in virtue of the fact that the additional person Quentin is better off in f3 than in f2. And it's silent on whether f3 is worse than f1 in virtue of the fact that Polly is better off in f1 and in f2 than in f3.

EC remains silent on a third question as well: how f2 compares against f3. It's silent on whether f2 is worse than f3 in virtue of the fact that there is an alternate future that makes things better for Quentin—namely, f3. And it's silent on whether f3 is worse than f2 in virtue of the fact that there is an alternate future that makes things better for Polly—namely, f2.

What do we want to say in response to those questions? Let's start with the question of how f2 compares against f3 since it may well seem—seem—that both what we want to say in response to that question is itself perfectly clear and what we want to say in response to that question will very nicely clarify what we shall want to say in response to the first two (how f2 compares against f1, and f3 compares against f1).

What is immediately striking about f2 and f3 is that the only difference between them is that the positions of Polly and Quinten have been reversed. And it may well seem to us any such merely reversing change in positions is morally innocuous—and that that's so regardless of what is going on in any further accessible future (e.g., f1). When two futures contain exactly the same populations and the only difference between them is a mere reversal of positions between one person and another, do we really need to know any more than that to decide that the one future is exactly as good as the other? Can *who* is advantaged, and *who* is disadvantaged, make a moral difference? Tradeoffs often must be made. And it may often be hard to see how those tradeoffs ought to be made. But surely when the tradeoff itself represents a merely reversing change, moral law is indifferent, indeed, impartial, between whether it's Quentin, or Polly, who sustains the loss of wellbeing.

Surely, in short, whatever else we want to say about this case, we shall want to say that f2 is exactly as good as f3.

The widely held and well known principle of anonymity supports that position.

Anonymity. For any futures x and y such that y is accessible relative to x, if exactly the same people do or will exist in x and y and each person in x has exactly the same wellbeing level that that same person has in y *except* for a *merely reversing change* across some those people, then x is exactly as good as y.

Anonymity easily secures the result that f2 is exactly as good as f3.

That result in hand, it may well seem that we are now very close to a complete account of the case. Now that we know that f2 is exactly as good as f3, we can appreciate just how much is at stake when we turn to compare each of f2 and f3 against f1: for we now know—"know"—that whatever we say about how *f*2 compares against f1, we must also say about how *f*3 compares against f1, and whatever we say about how *f*3 compares against f1.

So let's ask: do we say that both f2 and f3 are worse than f1—perhaps on the ground that both f2 and f3 include people who exist and suffer while f1 nicely avoids all such morally significant losses? Or do we instead say that both f2 and f3 are at least as good as f1?

To ask the question is practically to answer it. Whatever our instincts about how f3 compares against fl, it just seems churlish to insist that f2 is worse than fl—to insist that moral law *requires* that Quentin not be brought into an existence that is, though not maximized at +10, nonetheless at +9 *well* worth having.

We thus want to say that f2 surely *isn't* worse than f1.⁸ And—that claim in hand we now have a complete account of the case: f2 isn't worse than f1 (that is, f2 is at least as good as f1); f2 is exactly as good as f3; so f3 isn't worse than f1 (that is, f3 is at least as good as f1); and f1 isn't worse than f2 or f3 (that is, f1 is at least as good as f2 and f3). Appealing to various conceptual principles, we can then conclude the following: f1 is exactly as good as each of f2 and f3.

If *anti-natalism* is a position that generally prefers nonexistence over existence and that specifically declares both f2 and f3 worse than f1, then we can call the position that declares f2 and f3 at least as good as f1 anti-anti-natalist. Moreover, if *pro-natalism* generally prefers existence across the board and specifically declares f1 worse than both f2 and f3, then we can call the position that f2 and f3 are each, though at least as good as f1, not better than f1 anti-pro-natalism. Take your pick!

Now, there's nothing obviously problematic or particularly puzzling, as far as I can see, in the account we've just given of simple addition and reversal. It seems—*seems*—entirely plausible. But any confidence we may have in that account is soon to be shaken.

1.2 Indefinitely iterated addition and reversal

Thus consider the case we generate by simply reiterating the pattern that we see in simple addition and reversal along with the account we've just given of that case as many times as necessary to obtain a result that—we intuitively think—can't be

⁸ For the moment, we won't query exactly what principle supports the result that f2 isn't worse than f1 (that f2 is at least as good as f1). But we will come back to that question toward the end of this paper.

correct: a result that we find morally repugnant—every bit as repugnant as the idea that a future A in which a *very* large number of people enjoy lives *well* worth living isn't *better* than—and may even be *worse* than—an alternate future Z in which all those same people and many more besides have lives that are only *barely* worth living.⁹

Thus indefinitely iterated addition and reversal:¹⁰

choice	c1	c2	c3	c4	c5		cn
probability	1	1	1	1	1	1	1
wellbeing	f1	f2	f3	f4	f5		fn
+10	Polly	Polly	Quentin	Quentin	r		s
+9		Quentin	Polly	Polly	Quentin		
+8				r	Polly		
+.00003 +.00002 +.00001							Polly Quentin r
+0	Quentin,	r s	rs	S	S		
	r, s						

Figure 1.2 Indefinitely iterated addition and reversal

That the iterated line of reasoning generates a conclusion that is indeed repugnant becomes especially clear if we think of each of Polly, Quentin, r, s and etc. not as single person but as cohorts each consisting of billions or trillions of people (or more).

Now, it's true that there's a distinction to be drawn between Parfit's repugnant conclusion case and indefinitely iterated addition and reversal: in the new case, the existences just barely worth having come with still other existences that are (the case stipulates) well worth having.

But it's very unclear that focusing on that detail effectively sanitizes the result that fn is at least as good as fl. It's very unclear that the entirely avoidable suffering of billions or trillions of people can be made perfectly all right by the fact that that suffering has made it possible for billions or trillions more people to have the perfectly wonderful existences they in fact have.

⁹ Here, of course, I borrow from Parfit. See Parfit 1987, pp. 381–390.

¹⁰ See note 2 above.

And now we do have a puzzle: the *anonymity puzzle*. We accepted, in connection with simple addition and reversal, a line of reasoning that produced the result that f2 and f3, being exactly as good as each other, are also exactly as good as f1. We found that position—that anti-anti-natalist, anti-pro-natalist position—entirely plausible. We then reiterated the fact pattern, along with the account itself, indefinitely to produce indefinitely iterated addition and reversal. And we now find ourselves saddled with a result that seems highly implausible, indeed, repugnant. What has gone wrong?

2. When the cures are worse than the disease

2.1 Accept the repugnant conclusion?

Budolfson and Spears argue that many population theories face some version or another of Parfit's repugnant conclusion and thus that whether a theory generates such a conclusion—e.g., that Z is at least as good as A—can't be counted against that theory. An advantage of the average view over the total view is widely considered to be that the average view avoids the repugnant conclusion. Budolfson and Spears, however, argue that even the average view comes with its own version of the repugnant conclusion.¹¹ Once we see that the average view, just like the total view, generates the same old repugnant results, the average view is credibly (or at least arguably) left with no particular advantage over the total view.

Extending their point to the case at hand, we would simply say that the result that fn is at least as good as fl isn't, after all, a conclusion that we need to take pains to avoid. *All* theories generate some version or another of the same old repugnant result; and hence that any one theory generates such a result can't be counted *against* that theory.

The problem with that approach is that it comes with no demonstration that *all* population theories—or even that *all* facially *plausible* population theories—generate some version or another of the repugnant conclusion. *Perhaps* all population theories *based on the raw addition principle* (e.g., totalism) generate some version of the repugnant conclusion. But that's simply a reason to *question* the raw addition principle; it's not a reason to *accept* the raw addition principle. We don't try to browbeat ourselves into giving up the intuition that Z is worse than A. Nor should we try to browbeat ourselves into giving up the intuition that fn is worse than fl.

¹¹ Spears and Budolfson 2021. The position they describe is consistent with the view that the repugnant conclusion that Z is better than A really is repugnant. Their point, rather, is that if every credible theory generates that same repugnant conclusion, then we can't consider the fact that a theory generates that conclusion objectionable. Still other theorists have argued that the repugnant conclusion—whether as an objection to the total view or any other view—isn't, on closer inspection—truly "repugnant" at all. Huemer 2008.

Consider an analogy. Imagine a world in which most published theories of arithmetic instruct that 2 + 2 = 5. We don't then say that the fact that our theory implies that 2 + 2 = 5 can't be counted against our theory. We don't say, *oh well*, let's just set intuition aside; we don't try to browbeat ourselves into *believing* that 2 + 2 = 5. We instead *go on to the next theory*.

Deeply held, widely shared intuitions will be upended from time to time. But before abandoning an intuition of that sort—before, more precisely, *trying* to abandon an intuition of that sort—we need to have in hand a *new platform of deeply held, widely shared intuition*: one that loosens the hold the intuition we started with has on us; one that helps us understand exactly where the one intuition has gone wrong.

Similarly, before we accept that, in the end and after all, Z is as at least as good as A or fn is at least as good as f1—before we abandon our deeply held, widely share intuitions to the contrary—we need to have in hand an understanding that renders the old intuition no longer a genuine *intuition* at all.

Can we really be as confident that Z is worse than A or that fn is worse than fl as we are that 2 and 2 don't add up to 5? Well, can we be as confident that it's wrong to torture the dog for no reason whatsoever—that the future in which the dog is tortured is morally worse—as we are that 2 and 2 don't add up to 5? I think that we can be—and that the leap from the claim that the future in which the dog is tortured really is worse to the claim that Z and fn alike really are worse isn't a very impressive one.

2.2 Accept anti-natalism?

We earlier noted, and brushed aside as *churlish*, the position that f2 and f3 are both worse than f1. That wasn't a very well-articulated objection against anti-natalism. But to do any better job, we must first understand just why someone might accept that account.

An argument in favor of anti-natalism that at first glance might be considered plausible is this.¹² Let's take for granted that fn is worse than fl in indefinitely iterated addition and reversal. And let's then just note that it's plausible to think that the moral analysis that tells us (for indefinitely iterated addition and reversal) that *fn* is worse than fl will also tell us that (both for that same case and for simple addition and reversal) *f3* is worse than fl.¹³

¹² David Benatar is widely known as an advocate for anti-natalism. See Benatar 2006. This is not to suggest, however, that, faced with simple addition and reversal he would necessarily argue that f2 and f3 are both worse than f1. As far as I know, the case isn't one he's specifically considered.

¹³ A point of clarification: on my view, the "f1" that appears in simple addition and reversal and the "f1" that appears in the indefinitely iterated case are in fact distinct futures. (The same point holds of "f2"

What moral analysis might that be? What is the analysis that would confirm for us that f3 is worse than f1?

Here, a certain Pareto principle, adapted for use in the additional person context, seems highly plausible. Thus consider the *Pareto minus*, which, unlike its more standard ("same people") cousin, isn't restricted to the case in which the two futures to be compared contain exactly the same people:

Pareto minus principle. Where a future y is accessible relative to a future x and each person who does or will exist in y also does or will exist in x—where, that is, x contains each person y contains and may contain additional people as well—

x is morally worse than **y** *if*

there is a person p who does or will exist in x and x is worse for p than y and there is no person q who does or will exist in y such that y is worse for q than x.

According to this principle, when things can be made better for the *one* person simply by *not* bringing *another* person into existence to begin with, then, other things equal, it makes things worse to go ahead and bring that additional person into existence.

Since f1 makes things better for the existing Polly than f3 and since it does so by way of leaving Quentin out of existence altogether—and not by way of making things worse for anyone else—the Pareto minus principle immediately implies that f3 is worse than f1.

To conclude the argument to anti-natalism, then, all we need is a single application of the principle of anonymity: having determined that f^3 is worse than f1, we can now—thanks to anonymity—easily infer that f^2 is worse than f1 as well.

* * *

I can think of no more persuasive line of reasoning in favor of the anti-natalist account of simple addition and reversal than the one I've just laid out.

Moreover, it's a significant plus that the account tells us, not just that f3 is worse than f1, but also that fn is worse than f1.

But the account also comes with an obvious minus. We've now carefully worked through the most persuasive argument that we—that I—can think of to the con-

and "f3.") The future with respect to which fn, e.g., isn't accessible and the future with respect to which fn is accessible are, in other words, two distinct futures. That point of clarification has no practical import for purposes here. But see Roberts (manuscript).

clusion that f2 and f3 are both worse than f1. But that work hasn't—it seems to me loosened the hold one of the intuitions on which the puzzle itself is based has on us: the intuition that—whatever we want to say about f3-f2 really can't be worse than f1. That work hasn't even begun to dislodge our deeply held, and, I suspect, widely shared intuition that f2 is at least as good as f1. It hasn't presented us with a new platform of deeply held, widely shared intuition that helps us to understand why it is that f2 is, after all, worse than f1—understand, that is, exactly where our original intuition has gone wrong. *Churlish* still seems exactly the right word for the idea that f2 is morally worse than f1: that moral law *proscribes* Quentin's coming into existence.

But that means that the anti-natalist solution to the anonymity puzzle doesn't actually *solve* the puzzle at all but rather simply *tosses out* one of the puzzle pieces. We aren't going to have a *solution* to the puzzle until we have an account of the case that doesn't leave us intensely suspicious that it's not the intuition, but rather the argument against that intuition, that has gone wrong.

3. The inductive solution to the anonymity puzzle

3.1 Strategy

The solution to the anonymity puzzle I want to propose combines three ideas: that f3 is in fact worse than f1; that the position that f3 is worse than f1 *doesn't* commit us to the position that f2 is worse than f1; and that we can, in inductive fashion, build on the result that f3 is worse than f1 to show that f2 is, after all, at least as good as f1.

As we shall see in parts 3.2 and 3.3 below, the proposed solution is *inductive* in nature, in virtue of the fact that it relies on betterness results that we have already secured to generate betterness results that we are now in need of to complete a plausible account of the case.

That account, as we shall also see, involves a closer look at the anonymity principle itself. Thus the result that *f*3 is worse than f1 may well have never seemed off to us at all. Indeed, we may have recognized that we shall need to accept that *f*3 is worse than f1 *given that* we want also to say that *fn* is worse than f1.

We, accordingly-for the moment-take it for granted f3 is worse than f1.

What we *can't* take for granted—and not just, I think, for the moment—is that that *f*2 is worse than f1.

But the joint credibility of those claims—that f3 is worse than f1 and that f2 isn't worse than f1—shows clearly, if nothing else, that the principle of anonymity requires a closer look.

Anonymity considers f2 and f3 just two peas in a pod; how f2 compares against f1 dictates how f3 compares against f1 and vice versa. We need to consider whether, released from the constraints of anonymity, we find the door open *both* for the position that f3 (and fn) is worse than f1 *and* for the position that f2 *isn't* worse than f1.

3.2 Resurrecting the Pareto minus principle

As just noted, we seem easily able to agree that fn is worse than f1—and that the analysis that supports that result will also instruct that f3 is worse than f1.

But on what basis do we make those claims?

The obvious candidate is one that we have already introduced: the Pareto minus principle. According to that principle, since f1 makes things better for Polly than f3 without making things worse for anyone else—that is, for anyone who does or will exist in f1—f3 is worse than f1.

Now, it's true that that principle played a critical role in generating anti-natalism, and it's true that we've now rejected anti-natalism as a solution to the puzzle. But that doesn't mean that there's anything wrong with the principle itself at all.

Indeed that principle seems close to axiomatic—*assuming* that we really do think that existence is *just different*; that an *existing* child's suffering a terrible disease in a case in which that child's disease can accessibly be cured is just on a completely different moral plane from a possible child's never existing at all, notwithstanding the fact that *both* children have full moral status and that *both* children face the option of a zero wellbeing level, one as a result of the disease, the other as a result of never having existed at all.

Thus, the existence condition (EC) tells us that the worth having existence doesn't, other things equal, make things morally better: that it's perfectly permissible, other things equal, for agents not to bring additional people into existence. The Pareto minus principle takes things just one short step further: if, per EC, a person's *not* existing at all doesn't, other things equal, make things worse, and if, as we can all surely agree, a person's *existing and suffering* can *easily* make things worse, then when we have the choice whether to impose the one sort of loss on one person or the other sort of loss on another person, we'd—often—best go with the one sort of loss.

Applying that idea—and, specifically, the Pareto minus principle—we immediately infer that f3 is worse than f1.

Now, it would be a serious problem for the current strategy if that same principle also instructs that f^2 is worse than fl. But it doesn't. In view of the fact that f3 is better for Quentin, who exists in f2, than f2, the principle will remain silent on the question whether f2 is worse than fl.

3.3 Application of inductive principle; extension of the existence condition (EC)

So far, so good. We now turn to why it might be that, while f3 *is* worse than f1, f2 *isn't* worse than f1.

An approach that provides a basis for the claim that f2 isn't worse than f1 can be described as *inductive* in nature. It allows a given account of a given case in effect to build on itself, not in a way that is circular, but in a way that uses *already established* betterness results to reason our way to *further* betterness results.¹⁴

The idea, roughly, is just this. In effect, we have (in part 3.2 above) just provided grounds for the position that f3, though perfectly *accessible*, i.e., *within the bounds* of agents' ability, power and resources, is, as it were, *morally out of bounds*. The only options that are then left—the only options that are both accessible and *not* yet eliminated as morally out of bounds; as morally, that is, *worse than* still other futures—are just f1 and f2. Since f2 is better for Quentin than f1 and f2 is exactly as good for Polly as f1, we should have no problem at all declaring f2 at least as good as f1.

We can easily provide a rationale for that strategy. For purposes of this paper, it has been assumed that futures that aren't *accessible*—futures that lay outside of the agents' abilities, power, resources¹⁵—are irrelevant to determining whether a given future x is worse than a given future y. To explain the basis for that assumption, consider a certain *two outcome case*. In that case, either Charlotte shall never exist at all or Charlotte will exist and have a wellbeing level of +5. It won't work to try to make sense of that case by denying that there's a *logically possible* future in which Charlotte—at +10—is still *better* off than she is at +5. Of course there's such a future! But that fact isn't enough to make us worry that the future in which Charlotte is at +5 is—in the *two* outcome case—worse than the future in which She never exists at all. We remain confident in the claim that the future in which Charlotte is at +5 *isn't* worse than the future in which she never exists at all in view of the fact that the merely *logically possible* future—the future in which she is at +10—*wasn't* a future that agents (acting as individuals or together, whether collaboratively or not) had *any hope* of bringing about at all.

Thus the mere *logical possibility* of additional futures, however amazingly awesome they might be, wasn't enough to *silence* EC. Instead, EC immediately gene-

¹⁴ The inductive strategy we shall put to work here isn't inductive in the sense that it will use probabilistic, or statistical, reasoning to generate conclusions that are at best probably true. It's, rather, inductive in the sense that it fashions principles that don't just generate, in standard deductive form, all the results we're interested in in one fell swoop but rather allows the prior results generated by those principles to generate, by further application of those same principles, further results. Just to note: the new results generated follow by means of deductively valid reasoning—assuming the premises are true, so, we can be sure, are the conclusions.

¹⁵ See note 4 above.

rates, in the two outcome case, the result that the future in which Charlotte exists at +5 is exactly as good as the future in which she never exists at all.

We can now draw an analogy between a given future's being deemed *inaccessible* due to its status as a *merely* logically possible future and a future's being deemed *morally out of bounds* due to its having been already determined to be *morally worse* than an alternate accessible future by application of still other principles.

The idea—and here we go back to simple addition and reversal—is that the fact that f3 makes things better for Quentin shouldn't be enough to silence our analysis in the case where f3 has already been *ruled out as morally worse* than still another accessible future—that is, f1. f3 is, as it were, *morally* inaccessible in virtue of the fact that f3, per the Pareto minus principle, has already been shown to be worse than f1. We should thus be free to proceed to determine whether f2 is worse than f1 *as though f3 weren't an accessible alternate future at all.*

The fact that f3 is better than f2 for Quentin in simple addition and reversal shouldn't, in short, get in the way of the result that f2 isn't worse than f1 since *we've already shown* f3 to be, though perfectly *accessible, morally out of bounds*.

We can easily incorporate the inductive step into our solution by simply amending—in effect, extending—EC.

Existence condition with induction (ECI): Where x and y are possible futures and y is accessible relative to x,

x is morally worse than y, and a choice c made at x is wrong, *only if* there is a person p and an alternate accessible future z such that:

p does or will exist in x, and

x is worse for p than z (where z may, but need not, be identical to y), and

z isn't ruled out (by prior application of a sufficient condition on moral worseness) as morally worse than any alternate future also accessible relative to x.

Applied to simple addition and reversal, we can note, first, that no alternate accessible future exists that makes Polly better off than Polly is in f2. And then we can note that no alternate accessible future exists that makes Quentin better off than Quentin is in f2 that hasn't already been determined to be morally worse than still another future. It's true that f3 makes Quentin better off than f2. But f3 has *already* been ruled *morally out of bounds* by virtue of the fact that it's *already* been ruled *morally worse* than f1. We thus can ignore it, according to ECI, for purposes of determining whether f2 is worse than f1. Both of the people who do or will exist in f2 having been made as well off, either (in the case of Polly) as she accessibly can be made or (in the case of Quentin) as moral law itself allows, we easily conclude—and indeed ECI now implies—that f2 is at least as good as f1.

Now, in dealing with simple existence and reversal, the principle that gave us a starting point for our inductive analysis was the Pareto minus principle. (Once we secured the result that f3 is worse than f1, we could then move on to say that f2 is at least as good as f1.) But there's no reason to think that in other cases other principles might generate a starting point. Thus consider the following case:

choice	c1	c2	c3
probability	1	1	1
wellbeing	f1	f2	f3
+5	р		
+4		р	р
+3		q	r
+1	q		
+0			

Figure 3.3 Robin versus Sam

The existence condition in its original form—EC—will remain silent on whether f2 is at least as good as f3 (or vice versa). That's so, since there is in fact an alternate accessible future—f1—that makes things better for p than f2 does. But suppose that we also have in hand a priority principle we both find highly intuitive and have carefully tested. And suppose that principle implies that f1 is worse than f2. Then, the existence condition in its inductive form—ECI—by virtue of the fact that the priority principle has *already ruled* f1 morally *out of bounds* won't remain silent at all and will immediately tell us that f2 isn't worse than f3 and that f3 isn't worse than f2—that is, that f2 is exactly as good as f3.

3.4 More on anonymity; note on the mere addition principle

The problem with anonymity is that it whitewashes over the very distinction that we've just drawn between Quentin's situation in f2 and Polly's situation in f3. It *guts* the very project we have just engaged in: bringing to the surface the moral significance of the distinction between Quentin's situation in f3 and Polly's situation in f2.

Now, to reject the principle of anonymity *isn't* to reject any and all principles that *themselves* trade on the moral insight that anonymity itself tries to capture. Thus

nothing we've said here undermines the idea that certain other futures that differ only in respect of a merely reversing change are exactly as good as each other. When, e.g., two people simply reverse positions between two futures x and y, and the unique further accessible future z that is available to them is one in which neither person ever exists at all, we have no problem declaring x to be exactly as good as y. Nor does our rejection of Anonymity undermine a version of anonymity that has been restricted to cases involving futures all of which contain exactly the same people.

When anonymity is properly restricted, it's the epitome of fairness. We generally don't think moral law cares *who* you are. We think, rather, that moral law is, other things equal, *indifferent*, indeed, *impartial*, between the tradeoff being made in favor of the one person or in favor of the other person.

But we can accept those points while also insisting that the fact of a merely reversing change in additional people cases *doesn't* tell us *all that we need to know* about how the futures that constitute those cases are to be compared.

* * *

It might seem—*seem*—that a simpler way of securing the result that f2 is at least as good as f1 than going to the (minimal) trouble of deploying an inductive strategy and amending EC to reflect that strategy would be to appeal to the well-known and widely accepted *mere addition principle*. According to that principle, all we need to see, to infer that f2 is at least as good as f1, is that Quentin's existence in f2 is a *mere* addition: it's an existence worth having, and one that makes things worse for no one in f1 at all. How could such an addition make things *worse*?

But, on other grounds, that principle seems one we shall want to reject.¹⁶ And we've seen no reason in the context of the current discussion to think that that was a mistake.

And, as we've now seen, nor do we need that principle in order to solve the anonymity puzzle: we don't need that principle, or any principle, that begs the very questions we are now working so hard now to settle.

3.5 Summing up

The solution we've sketched here—the *inductive solution*; shorthand for induction on the fact that existence is, in the ways specified above, *just different*—rejects the

¹⁶ See Roberts (manuscript), Chapter 4 ("The Pareto Puzzle"). Among other plusses, setting aside the mere addition principle opens the door to an elegant solution to Parfit's own mere addition paradox. See Parfit 1987, pp. 419-441.

principle of anonymity in its unrestricted form. We haven't *flatly* rejected that principle. We haven't simply pointed out the obvious—that is, that we can save our intuition that fn is worse than fl by rejecting anonymity. Rather, we have articulated just *why* that principle, in its unrestricted form, plausibly *fails*: it fails in virtue of the fact that it *obliterates* distinctions that themselves ground an account of the case that asserts both that fn is, indeed, worse than fl and that f2 is at least as good as fl. Thus we solve the puzzle *without* throwing out any of the puzzle pieces.

4. Objections and replies

4.1 Does the inductive solution imply moral actualism?

In drawing the distinction between Polly's situation in f2 and Quentin's situation in f3, it might seem that we assign a certain moral status to one class of people we deny to others. Specifically, it might seem that we are assigning a moral status to Polly on the ground that Polly *exists* in *all three futures* and assigning no moral status to Quentin at all on the ground that Quentin *doesn't* exist in *all three futures*. We are holding it against Quentin that Quentin's existence is—we might say—merely *contingent*, and counting it in favor of Polly that Polly exists—we might say—*necessarily*. We are saying that *Quentin's* loss in f2 *doesn't* matter morally while Polly's loss in f3 *does* matter morally. But making a person's *moral* status a function of that person's *existential* status is a widely recognized mistake. We should, accordingly, reject moral actualism in all of its currently articulated forms.

In fact, however, the inductive solution *doesn't* rely on the idea that some people do matter morally and others don't. If we revised simple addition and reversal to include a fourth future, one in which neither Polly nor Quentin ever exists at all, we would obtain exactly the same results we've already registered: that f3 is worse than f1, and that f2 is at least as good as f1. Nor have we even noted, in either presenting or analyzing the case, *which* future happens to be actual; nor do we try to relativize our betterness rankings to particular subclasses of people (e.g., those who exist in f2 and f3 but not in f1). We are safe, in other words, from charges of moral actualism.

4.2 Isn't anonymity itself a critical piece of the puzzle?

Certainly, theorists for whom an unrestricted principle of anonymity is ingrained who feel certain, if of nothing else, that f2 is exactly as good as f3 in simple addition and reversal—will think that that principle is a critical piece of the puzzle. But those same theorists are accustomed to a framework in which the comparison of one future against a second can be accomplished without a glance at any third future. Which is just to say that they are accustomed to a framework in which certain details of the one future or the other future—including the details that distinguish Polly's position in f3 from Quentin's position in f2—are screened out of the analysis. Once we appreciate that those details aren't incidental at all and are instead critical to the analysis of the case, we can also appreciate that anonymity, though generally exactly right, can in the odd case get things exactly wrong.

One final note. Theorists for whom anonymity in its unrestricted form is ingrained also happen often to be theorists in which the repugnant conclusion in any of its various instantiations is also no big deal. (Either the conclusion isn't repugnant at all, or it's repugnant but, since it's an implication of all credible population theories, it's a conclusion we are forced to hold our noses and accept.) But it's worth underlining that a "solution" to the anonymity puzzle that simply accepts the result that fn is at least as good as f1 never can constitute an actual *solution* to that puzzle: simply denying that there ever was ever a puzzle to begin with—simply *throwing out* some of the puzzle pieces—is very different from offering an actual *solution* to the puzzle.

References

Benatar, David 2006. *Better Never to Have Been: The Harm of Coming into Existence*. Oxford: Oxford. University Press.

Budolfson, Mark and Dean Spears 2021. "Repugnant Conclusions." *Social Choice and Welfare*: 1–22.

Narveson, Jan 1976. "Moral Problems of Population." In Michael D. Bayles (ed.), *Ethics and Population*. Cambridge, Mass.: Schenkman, pp. 59–80.

Parfit, Derek 1987. *Reasons and Persons*. Oxford: Oxford University Press (originally published 1984).

Roberts, M.A. (manuscript). The Existence Puzzles.

Henrik Andersson & Anders Herlitz¹

Classifying Comparability Problems in a Way that Matters²

How should one understand cases in which neither of two alternatives is at least as good as the other? A major motivation for studying this is that failure to establish that some option is at least as good as every alternative pose a problem for practical reasoning. Meanwhile, the recent literature on comparability problems has almost exclusively focused on what the appropriate *explanation* of the failure to determine that an option is at least as good as all alternatives is, e.g. vagueness or the possibility of nonconventional comparative relations. This paper argues that the focus on how to best explain comparability problems has reached an impasse at which it is hard to make any progress, and realigns the discussion on comparability problems with what motivates an interest in them by introducing a new classification of these problems.

 $^{^1} Institute for future studies, and ers, herlitz@iffs.se, henrik.and ersson@fil.lu.se.$

 $^{^2}$ We gratefully acknowledge the financial support from Riksbankens Jubileumsfond (grant number grant number M17-0372:1).

How should one understand cases in which neither of two alternatives is at least as good as the other? A major (and often cited) motivation for studying this issue – and comparability problems in general – is that failure to establish that some option is at least as good as every alternative pose a problem for practical reasoning. In fact, some have even argued that it challenges the very possibility of rational choice (e.g. Hsieh 2007; Chang 2013). Meanwhile, the literature on comparability problems has almost exclusively focused on what the appropriate *explanation* of the failure to determine that an option is at least as good as all alternatives is, e.g. vagueness or the possibility of non-conventional comparability problems, first, is misguided since it has little relevance for what the implications for practical reasoning are, and, second, it has reached an impasse at which it is hard to make any progress. Instead, we argue that progress in this area can be made by focusing on comparability problems' implications for practical reasoning are, and, second, it progress in this area can be made by focusing on comparability problems' implications for practical reasoning and introduce a new classification of comparability problems that can facilitate progress.

Interesting comparability problems arise when it seems meaningful to compare two alternatives with respect to their value (or choice- or belief-worthiness) but yet we find it hard to state how they actually relate with respect to their value (or choiceor belief-worthiness). These cases are often referred to as *hard cases*. Here are two examples:

Career: A career as a lawyer and a career as an academic are both attractive and meaningful careers. However, the two careers will differ in such drastic ways so that it is difficult to determine which is better.

Creativity: Mozart and Michelangelo are both creative geniuses. However, their creativity seems to be so different in kind so that it is impossible to determine who is the most creative *tout court.*³

For a long time, hard cases were believed to be cases in which the 'at least as good as' relation did not apply. On this view, we can only say something about how the alternatives are *not* related, i.e., they are not related by the 'at least as good' relation (e.g., Berlin 1968; Williams 1981; Raz 1986). Lately, however, many philosophers have challenged this idea and it has been argued that even in hard cases a positive value relation apply. Two competing explanations stand out.⁴ On one view, hard cases can

³ Chang 2002: 659.

⁴These accounts are not in a strict sense competing, but adherents of the views treat them as such. We will have reason to return to this later on.

be explained in terms of vagueness (see Broome 1997, 2004; Wasserman 2004; Constantinescu 2012; Elson 2017; Andersson 2017). On this view, it is merely indeterminate *which* of the conventional comparatives 'better than', 'worse than' or 'equally as good' obtain. On another view, hard cases should be understood as cases in which none of the conventional comparative relations apply but another positive comparative relation obtains (see Parfit 1984, 2016; Griffin 1986; Chang 1997, 2002, 2005, 2016; Rabinowicz 2008).⁵

In this paper, we will argue that the two views do not differ as much as previously thought and suggest that insofar as one is interested in the challenges comparability problems pose for practical reasoning and rational choice, it is more fruitful to reconceptualize what the controversy is about. We argue that a new classification of comparability problems is needed and we suggest that a plausible and pragmatic classification should be grounded in what the implications for practical reasoning and rational choice are. Focusing on this is will be a fruitful path to progress in this area.

The paper is structured in the following way. In the first section, we introduce the *Non-Conventional Comparative Relations View* and the *Vagueness View* in more detail, identify the core points of disagreements, and show that the dispute boils down to a disagreement regarding what one takes to be the necessary properties and implications of vague predicates, something which has nothing to do with practical reasoning and rational choice. In the second section, we argue that a new classifycation of comparability problems that better corresponds to the practical challenges posed by them is needed. In the third section, the benefits of such a classification are highlighted. The fourth section concludes the paper.

1.

On one prominent view, comparability problems should be understood with reference to non-conventional comparative relations. Inspired by Derek Parfit's remarks in *Reasons and Persons* as well as by James Griffin's suggestion that items sometimes are 'roughly equal', Ruth Chang has famously argued for the possibility of a nonconventional comparative relation: 'parity' (Parfit 1984; Griffin 1986; Chang 1997,

⁵ We take it that most philosophers who write on the topic acknowledge that some hard cases can be due to ignorance. The disagreement should consequently be framed in terms of how we are to account for hard cases that cannot be given a satisfactory explanation in terms of ignorance. It should also be acknowledged that it is possible to hold the view that *all* hard cases ought to be understood in terms of ignorance. On this view, one object is always at least as good as the other; however, due to ignorance we sometimes cannot determine how they relate. This view has to our knowledge only been defended by Donald Regan who admits to taking on the self-proclaimed role of the 'designated eccentric' when arguing for this position (Regan 1997). So-called epistemicist accounts of vagueness could also be understood to entail such a position (cf. Williamson 1994).

2002). Parity, as Chang understands it, is a positive comparative relation that is symmetric, irreflexive and non-transitive. It resembles the relation 'equally as good as' in that neither of the items that are on a par is worse than the other, but it is distinct from 'equally as good as' in that it is irreflexive and non-transitive (cf. Carlsson 2010). On the view that parity is possible, the reason a career in law and a career in academia cannot be compared in terms of *conventional* comparative relations is that they bear a non-conventional comparative relation to each other. The two careers are on a par. Several philosophers have picked up on Chang's idea and defended it in various contexts and in various ways (Fröding & Peterson 2007; Carlson 2010; Schoenfield 2014). Others have argued for the possibility of other non-conventional comparative relations. Wlodek Rabinowicz has, for instance, illustrated that if one applies the fitting attitudes-approach to value, one can, on a conceptual level, identify no less than 12 possible, non-conventional positive comparative relations (Rabinowicz 2008). And Parfit has argued for the possibility of what he calls 'imprecise equality', which is different in kind from 'equally as good as' (Parfit 2016).

Another prominent view is that comparability problems should be understood in terms of vagueness (see Broome 1997, 2004; Wasserman 2004; Elson 2017; Andersson 2017). To see how vagueness can explain why it is sometimes the case that no alternative can be determined to be at least as good as the other, it is helpful to consider how all comparisons are triadic, and that when one makes a comparison of two items one always compares them with respect to something. For example, in the comparison of Mozart and Michelangelo above we are comparing them with respect to creativity. Following Chang, we can call this something with respect to which one makes a comparison a 'covering consideration' (Chang 1997). Although some covering considerations are very precise (e.g. age, price, length), certain covering considerations seem to be vague (e.g. 'balder'). Proponents of the Vagueness View hold that besides the hard cases that can be explained with reference to lack of knowledge (e.g. it is for epistemic reasons hard to determine whether Socrates' father died at a younger age than Thales' father) hard cases should be explained with reference to how covering considerations are vague. The reason a career in law and a career in academia are hard to compare on this view is that the covering consideration (e.g. 'goodness of career') is vague. When covering considerations are vague, it is not true that neither of the conventional comparative relations holds, but it is indeterminate which one holds. Furthermore, most adherents of the Vagueness View adopt a supervaluationist theory of vagueness, which holds that a disjunction may be true while none of its disjuncts are true (see Fine 1973; Broome 2009). Consequently, the disjunction 'A is better than, worse than, or equally as good as B' may be true while each disjunct is neither true nor false. While the Vagueness View is compatible with the existence of non-conventional value relations at a conceptual

level, proponents of the Vagueness View argue that the three conventional value relations and the phenomenon of vagueness is sufficient in order to account for all value comparisons. Proponents of the Vagueness View thus claim that there is no reason to invoke non-conventional comparative relations and that hard cases can be explained in terms of vagueness.

On a conceptual level it is easy to distinguish between the two views. For instance, if two objects are on a par it is determinately false that one of them is at least as good as the other, something proponents of the Vagueness View refute. However, both views aim at explaining more than conceptual possibilities. They aim at providing us with reasons why one conceptualization better explains hard cases than the other. It is here, in providing an explanation of the hard cases, that the views become hard to differentiate.

Much of the debate between proponents of the Non-Conventional Comparative Relations View and proponents of the Vagueness View has circulated around how we should interpret what Joseph Raz referred to as the 'mark of incommensurability':

We have here a simple way of determining whether two options are incommensurate given that it is known that neither is better than the other. If it is possible for one of them to be improved without thereby becoming better than the other or if there can be another option which is better than the one but not better than the other then the two original options are incommensurate (Raz 1986: 121).

For Raz, the mark of incommensurability can be used to determine whether two alternatives are 'incommensurate', i.e. whether the two alternatives are not determinately related by a 'better than', 'worse than', or 'equally as good' relation. That is, if it turns out that for two alternatives, A and B, A is neither better nor worse than B and a slightly improved A is not better than B then A and B are not determinately related by an 'at least as good as' relation. This test is also employed in the much-discussed Small Improvement Argument (cf. Chang 2002; Gustafsson 2013). The argument has the following structure:

Premise 1: It is false that A is better than B and it is false that B is better than A.

Premise 2: A+ is better than A.

Premise 3: A+ is not better than B.

Conclusion: It is false that A is better than B, it is false that B is better than A, and it is false that they are equally as good

Premise 2 and Premise 3 establish that A and B are not equally as good. 'Equally as good' is a transitive relation and, consequently, if A and B where equally as good, then, if A+ is better than A, A+ must also be better than B.⁶ From this and Premise 1 the conclusion follows. If Conclusion is correct, then the Vagueness View is false. The Vagueness View states that for any two objects, A and B, it is always true that A is better than B, B is better than A or A and B are equally as good (although it might be indeterminate which of these relations obtain). Conclusion, however, states the opposite.

The above characterization of the Small Improvement Argument presents the structure of the argument, but to rule out the Vagueness View it must also be shown that there are objects that stand in a relation that satisfies Premises 1, 2 and 3.⁷ Several examples of such objects have been given by adherents of the Non-Conventional Comparative Relations View.⁸ But in all these cases, it may be only indeterminate how the objects relate to each other (Gustafsson 2013; Rabinowicz 2009). For the Small Improvement Argument to undermine the Vagueness View, it must be shown that it is *determinately* false that A is better than B, and determinately false that B is better than A. Furthermore, it must be determinately true that A+ is better than A. It is hard to establish conclusively that three items relate to each other in such ways. For all of the alleged cases of non-conventional comparative relations, the adherents of the Vagueness View can counter and say that it is in fact indeterminate how the items relate, and precisely because of this, a small improvement to one of them will not make it determinately better than the other. So even if it is determinate that A+ is better than A, it may be indeterminate whether A+ is better than B and indeterminate whether A is equally as good as B. Thus, in order to provide

⁶ Reuter and Messerli (2017) have suggested that there is no need to assume transitivity. Their argument is a *reductio* that involves premises 1-3 and a fourth premise: A=B. From this they claim that one can substitute B in Premise 3 with A. That is, from substituting B with A in Premise 3 they derive that A+ is not better than A which contradicts Premise 2, and from this they conclude that it is not the case that A=B. However, they seem to still implicitly assume that 'equally as good' is transitive. Without this assumption the substitution cannot be made.

⁷ Furthermore, for the Non-conventional Comparative Relations View to be convincing, it must be shown that A and B can be comparable, even though they are not related by the standard relations. This possibility is argued for by Chang 2002. She employs the *Chaining Argument* in order to establish that alternatives that has the mark of incommensurability can in fact still be comparable. This is a truly interesting argument but we will not discuss the details of it since it is not of relevance for this paper.

⁸ For instance, how do Mozart and Michelangelo relate to each other with respect to creativity? It seems false that Mozart is better than Michelangelo and it seems false that Michelangelo is better than Mozart. We can, moreover, easily imagine a slightly improved Mozart, Mozart+, who, for instance, has produced one more sonata than Mozart. Mozart+ is better than Mozart with respect to creativity. Yet, most would hesitate to say that Mozart+ is better than Michelangelo. If this is true then Mozart and Michelangelo cannot be equally as good, since in that case the small improvement to Mozart would also have made him better than Michelangelo. Consequently, Mozart and Michelangelo are not related by any of the conventional comparative relations.

a convincing argument for the Non-Conventional Comparative Relations View it must be shown that vagueness does not explain hard cases.

Chang presents two such arguments. Chang's first argument focuses on the phenomenology of the cases. Paradigm cases of vagueness are co-called borderline cases (e.g. 'is it a heap?', 'is he bald?'). These cases, Chang claims, differ from cases in which non-conventional comparative relations have been suggested to obtain in the following way. In borderline cases, we tend to be as willing to say that a predicate applies as we are to say that it does not apply (Chang 1997: 682).⁹ For example, if Harry is borderline bald, we are just as willing to call him bald as we are to call him not bald. By contrast, in the cases where Chang thinks that alternatives bear some non-conventional comparative relation to each other we might be willing to say that neither option is better than the other (e.g., Mozart is not better than Michelangelo (with respect to creativity)), but we are *not* as willing to say that one option is better than the other (e.g., that Mozart is better than Michelangelo (with respect to creativity)).

Adherents of the Vagueness View are unimpressed by this argument. They object that the reason these two cases are different is not that one is a case of non-conventional comparative relations and one is a case of vagueness. There are other differences. The most obvious difference is that one case involves a comparative while standard borderline cases involve monadic predicates (cf. Wasserman 2004). Although most paradigmatic examples of vagueness concern monadic predicates, also comparatives can be vague. Famous examples include comparatives that are vague due to multidimensionality (see Keefe 2000: 14). Luke Elson writes:

In *comparative* borderline cases, the relevant question is [...] 'is a Fer than b?' If Hank has fewer hairs widely distributed over his head, and Henry has more thick hairs concentrated in a ring around his scalp, then it may be indeterminate or unknowable whether Hank is balder, or Henry is balder, or they are precisely equally bald (Elson (2014: 7).

⁹ Johan E. Gustafsson has presented an interesting objection to this claim. He notes that even if there is vagueness it need not be the case that we are as willing to judge that a predicate applies as we are willing to judge that it does not apply. In his example, there are two borderline bald men have very similar hair, but one of them, Harry, is more hairy than the other, Larry. "Even though both are borderline cases of baldness, we might be less willing to call Harry more bald than Larry. Yet we would not therefore be less willing to call Harry not bald. Thus the extent to which one is willing to judge that a term applies in a borderline case can be lesser than the extent to which one is willing to judge that it does not apply Gustafsson" (2011: 441).

Following this idea, we can compare the case of Mozart and Michelangelo with a case of comparative vagueness.¹⁰ Assume that we are to determine who of two borderline bald men, Harry and Curly, is balder. Harry's and Curly's distributions of hair are different; Harry has more hairs, but these are patchy; Curly has fewer hairs, but these are more evenly spread out. What are our intuitions about a case like this? If we are willing to say that Harry is not balder than Curly, but not willing to say that Harry is balder than Curly, but not willing to say that Harry is balder than Curly, but not willing to say that Harry is balder than Curly are between this case and cases where non-conventional comparative relations are supposed to obtain. Those who support the Vagueness View argue that Chang's first argument gives us no reason to not understand hard cases of comparisons in terms of vagueness.¹¹

Perhaps envisioning resistance, Chang provides a second argument for the Nonconventional Comparative Relations View:

Perhaps the force of the argument from phenomenology is not altogether clear. In that case, we might allow that there is some "perplexity" over whether one item is better than the other, where this perplexity is consistent with the possibility of semantic indeterminacy. The question then is whether this perplexity has its source in the vagueness of predicates. This question can, I believe, be answered by examining the way this perplexity might be resolved and comparing it to the way borderline cases of a vague predicate might be "resolved." Of course in one way, there is already a "resolution" in a borderline case: it is borderline. But there is a perfectly clear sense in which we can nevertheless ask, How are we to resolve its borderline status? That is, we ask, apart from any context, the following hypothetical: If we had to choose between application or not, how would we do so—what would be a permissible way of resolving the indeterminacy? It is in this broad, intuitive sense of "resolution" that we can ask whether the resolution of perplexity in [...] cases [where non-conventional comparative relations obtain] is like the resolution of indeterminacy in borderline cases (Chang 2002: 682).

According to Chang, the *resolution* in these cases differs. When there is vagueness, she suggests, it is permissible to resolve the indeterminacy by 'arbitrary stipulation' (Chang 1997: 682). That is, when someone is borderline bald, it is permissible to use an arbitrary method (perhaps flip a coin) to determine whether the person should be referred to as being bald or not. This does not mean that the question of whether the person is bald or not can be determined by such arbitrary procedure; it is still

¹⁰Wasserman's example is similar to the one presented here. "Suppose that Harry has 100 hairs distributed more-or-less evenly across his scalp. Suppose that Curly has 99 hairs that are perfectly distributed across his scalp. Is Harry balder than Curly?" Wasserman (2004: 396).

¹¹ Wasserman 2004; Elson 2014; Andersson 2017.

indeterminate whether he his bald or not, but it is permissible to resolve the indeterminacy arbitrarily. Contrary to this, there are hard cases in which it is not permissible to arbitrarily determine a conventional ranking according to Chang. Say that one is to compare Mozart and Michelangelo with respect to creativity. In such a case, it seems impermissible to flip a coin in order to determine whether Mozart should be referred to as the better artist with respect to creativity. Moreover, if two people do flip a coin and end up with different rankings of Mozart and Michelangelo this would be a "substantive disagreement in which arguments can be brought to bear" (Chang 2002: 685). The same cannot be said if they had disagreed about whether a borderline bald person is bald. In the latter case, there would, on Chang's view, be no substantive disagreement but only "a clash of arbitrary decisions in the face of indeterminate application" (Chang 2002: 684). In the case where Mozart and Michelangelo are compared with respect to creativity, there is a *resolutional remainder*. When one assesses whether a borderline bald man is bald or not, there is no such resolutional remainder.

Indeed, it seems almost essential to hard cases that we cannot just make an arbitrary choice between the alternatives. It is hard to refute the following claim by Chang:

Consider a superhard case involving comparison of a particular act of promise keeping and a particular act of bringing about great happiness with respect to moral goodness. As the case is hard, the promise keeping is morally better in some respects—for example, it fulfills one's obligation to keep promises—while the bringing of great happiness is better in other respects—for example, it addresses legitimate interests of many people—and yet it is not obvious that one is morally better than the other overall. Now the question before us is, could the resolution of the case be an arbitrary matter—could the perplexity concerning which is morally better be answered by the flip of a coin? Clearly, the resolution of this superhard case cannot be a matter of arbitrary stipulation but is a substantive matter concerning which is better (Chang 2002: 685).

It is, however, questionable whether the existence of resolutional remainders settles the issue of whether we ought to accept the Non-Conventional Comparative Relations View or the Vagueness View. There is nothing that commits proponents of the Vagueness View to the idea that we can arbitrarily stipulate how alternatives relate. That is, one can allow for the existence of resolutional remainders within the framework of the Vagueness View.

In fact, there are several possible ways to make room for resolutional remainders when dealing with vague comparatives. For instance, one could claim that the consequences of a choice actualize resolutional remainders, and that this is unrelated to how one explains comparability problems. This becomes very clear when considering comparatives such as "morally better than". When making moral comparisons we ought to treat the topic with a certain amount of respect and flipping a coin seems to be too brute an act to determine what to do. But this would be true for paradigmatic examples of vagueness too. If labelling Harry as balder than Curly has significant consequences, it would seem wrong to determine the labelling with the flip of a coin. This is just one example of how one could go about to argue for the claim that there is a resolutional remainder even when it comes to vague comparatives.

Adherents of the Non-conventional Comparative Relations View might reply that it is not *essential* to vagueness that it gives rise to resolutional remainders, while it is essential in the hard cases they discuss. However, this is rarely how non-conventional comparative relations are defined or how proponents of this view suggest that we understand them, and it is not obvious that proponents of these explanations want to take this route. If resolutional remainders are an essential feature of choice situations in which some alternatives bear a non-conventional comparative relation to each other, then the above-mentioned perplexity ought to appear in all situations in which these relations are supposed to appear. This goes against our experiences. To give an example, there is no reason to assume that non-conventional value relations cannot obtain in insignificant everyday situations. Perhaps the comparison of an apple and a pear in terms of which is the better afternoon snack could be believed to be a case of e.g., parity. It seems odd to maintain that there must be a resolutional remainder in that case.¹²

Moreover, note that it is not sufficient for the proponents of non-conventional comparative relations to claim that it is not essential for vagueness to give rise to resolutional remainders. That claim could be compatible with some cases of vagueness giving rise to resolutional remainders. Proponents of the Non-Conventional Comparative Relations View must make the stronger claim that it is essential for vagueness that it never gives rise to resolutional remainders.

The whole debate between the two camps seems to boil down to whether vagueness has some property that it is typically never associated with or not (i.e. a property which justifies arbitrary resolutions). In fact, Chang admits that it is a genuine possibility that vagueness does not always have this property, i.e. that there might

¹² It could in fact even be questioned whether there is a resolutional remainder in cases such as the comparison of Mozart and Michelangelo with respect to creativity as well. If one weighs in the consequences of labelling Mozart better than Michelangelo then there might be such a remainder, but if there are no consequences it seems less obvious. Here our intuitions and adherents of non-conventional comparative relations intuitions might, however, come apart.

be a kind of vagueness "whose borderline cases are always open to substantive argument. But if there is, it remains to be explained" (Chang 2002: 686). The question then is why one should think it is essential to vagueness that it does not give rise to resolutional remainders. Three features are often mentioned to be essential for vague predicates: they admit borderline cases, they lack sharp boundaries, and they are susceptible to sorites paradoxes (Keefe 2000: 6). The lack of resolutional remainders or features that might resemble these are generally not mentioned in the discussions.¹³ Furthermore, if labelling Harry as balder than Curly has significant consequences, then it is natural to assume that there would be resolutional remainders if we determine the labelling with the flip of a coin.

From this we conclude that the claim that it is essential for vague predicates that there is no resolutional remainder cannot be taken to be a generally accepted feature of vague predicates. In fact, hard cases could even be construed as illustrations that support the claim that there can be a resolutional remainder when we have a case of comparative vagueness.

More interestingly, the whole debate around whether hard cases should be understood in terms of vagueness or in terms of non-conventional comparative relations seems to take a halt at the discussion concerning resolutional remainders.¹⁴ In fact, at this point the discussion seems to boil down to the issue of what is essential for vagueness. If we define vagueness such that it never gives rise to resolutional remainders, then hard cases are not cases of vagueness but rather cases of nonconventional comparative relation. If we, on the other hand, define it such that it allows for resolutional remainders, then hard cases could be understood in terms of vagueness. This is a surprising result. The whole debate has now been reduced to a question of how we define vagueness. On one account of vagueness, an argument can

¹³ The idea could find some support in the subvaluationist theory of vagueness (cf. Hyde 1997). According to this theory, a predicate is true if and only if it is true on some precification, and a proposition is false if and only if it is false on some precification. This means that for borderline bald Harry, it is both true and false that he is bald. Vague predicates can, consequently, be both true and false. If this theory of vagueness is correct, it could be argued that it is essential for vague predicates that there is no resolutional remainder. For subvaluationism, it is essential that there is no room for substantive disagreement when one person claims that Harry is bald and another claims that he is not bald, since it is both true and false that he is bald. Subvaluationism, however, is heavily disputed and not generally accepted. This in itself can be taken to be a token of the fact that resolutional remainders are not essential features of vagueness; some do indeed find it odd that it is both true and false that Harry is bald. To them it cannot be the case that vague predicates have this feature.

¹⁴ It should be acknowledged that some adherents of the Vagueness View believe that the debate can be settled by appealing to the 'Collapsing Argument'. We will not go in to details and describe this technical argument but suffice to say that by accepting a highly controversial principle, 'the Collapsing Principle', it can be showed that vagueness leaves no room for non-conventional value relations (Broome 1997). That is, if we accept that most reasonable claim that comparatives can be vague then the Non-Conventional Relations View must be rejected. Even if some defend the contested Collapsing Principle (Constantinescu 2012; Andersson & Herlitz 2018) many find the principle questionable and thus the argument seems not to have settled the debate.

be made that hard cases are cases of vagueness, and on another account of vagueness, an argument can be made that hard cases are cases of non-conventional comparative relations. However, as we will show there is no need to focus on these technical details in order to make progress what matters when making a rational decision does not seem to depend on these details.¹⁵

Nevertheless, the debate has taught us a great deal. One interesting result might be that non-conventional comparative relations might be found in a bigger domain than previously thought. They might appear in all cases we standardly refer to as cases of multidimensional vagueness, and not only in the evaluative realm. This seems like a natural conclusion for proponents of the Non-Conventional Comparative Relations View. Why expect there to only be evaluative parity when the formal features that grounds parity also can be found within the non-evaluative realm? Another interesting result is that hard cases seem to come in two shapes. There are hard cases that actualize resolutional remainders (or something like them), and hard cases that do not. We turn to the importance of this result now.

2.

The debates between proponents of the Vagueness View and proponents of the Non-Conventional Comparative Relations View reveal that there are differences between different kinds of hard cases. Notably, some hard cases actualize a need for further reasons to determine what to do while other hard cases do not. In this section, we argue that a new classification of hard cases ought to be introduced. A distinction that invites itself naturally is considered. The upshots and possible drawbacks of this distinction is considered and the desideratum for a classification of hard cases can be presented. The aim is to establish that regardless of who is right about the question of whether hard cases should be understood in terms of vagueness or in terms of non-conventional comparative relations, this new classification provides tools for how to think about how to form reasonable decisions when no option is determinately at least as good as every alternative.

In order to discuss possible classifications, let us first introduce some additional terminology. Call an option *optimal with respect to F* if it is determinately at least as F as every alternative. Call an option *maximal with respect to F* if it is *not determinately* less F than any alternative.¹⁶ "F" denotes the relevant consideration of the

¹⁵We are not the only ones to point out the possibility of non-conventional comparative relations being the same as vagueness. Johan Gustafsson presents an interesting argument in which he argues that 'parity' ought to be construed in terms of vagueness. (Gustafsson 2013).

¹⁶ It is worth noting that maximality is typically defined so that an alternative is maximal if it is not worse than any alternative (see, e.g., Sen 1997), and that our definition thus is more precise in that it refers to what is fully determined to be the case.

choice situation. For example, when comparing two possible holiday destinations "F" can be substituted by "priceworthy holiday destination", "adventurous holiday destination", "relaxing holiday destination" or some combination of several important considerations. The meaning of "F" is thus determined by the agent when considering the choice situation.

The kind of hard cases that much of the debates on how to explain hard cases focus on are characterized by the fact that there is no alternative that is optimal with respect to what matters, while there are several alternatives that are maximal with respect to what matters. Neither Harry nor Curly is optimal with respect to baldness, but both are maximal with respect to baldness (since maximality here is defined in terms of being 'not determinately worse than any alternative'). Neither Mozart nor Michelangelo is optimal with respect to creativity, but both are maximal with respect to creativity. Proponents of the Non-Conventional Comparative Relations View and proponents of the Vagueness View disagree about how to explain that no alternative is optimal with respect to what matters while several alternatives are maximal with respect to what matters, but they agree on the general characterization that there is no optimal alternative, yet there are several maximal alternatives.

In light of this, we can describe the phenomenon that gives rise to the controversies around hard cases in a new way (cf. Herlitz 2019, 2020):

Nondeterminacy: x and y are *nondeterminate in their ranking* with respect to F if it is not determined that x is more F than y, not determined that x is less F than y, and not determined that x and y are equally as F.¹⁷

Nondeterminacy poses a practical problem if *x* and *y* are nondeterminate in their ranking and it is also the case that there is no alternative, *z*, that is determinately better than both *x* and *y* (i.e. *x* and *y* are both maximal with respect to *F* but neither *x* nor *y* is optimal with respect to *F*).

The discussion above made two things clear. First, there are different metaphysical explanations that can account for nondeterminacy. Second, the metaphysical explanation has no implications for what kind of problems nondeterminacy generates for practical reasoning.

A natural classification of nondeterminacy that is guided by an interest in understanding its implications for practical reasoning abandons the attempts to distinguish nondeterminacy due to vagueness and nondeterminacy due to non-con-

¹⁷ It should be clarified that when x is nondeterminate to y, and vice versa, it doesn't follow that x and y are nondeterminate to all objects in the domain.
ventional comparative relations. Instead, such a classification is based on the practical implications of the comparison. There are several reasons to favor a classification based on the practical implications over a classification based on metaphysical explanation. First, as argued above, the classifications that are based on what grounds nondeterminacy are at an impasse: distinguishing vagueness from non-conventional comparative relations have no obvious benefits since all the properties that might be associated with nondeterminacy due to vagueness can be associated with nondeterminacy due to non-conventional comparative relations, and vice versa. Second, nondeterminacy constitutes a significant potential practical problem: What can we reasonably do when some maximal alternatives are nondeterminate in their ranking? Classifying hard cases based on their practical implications facilitates research on this question since it connects the question of how to act reasonably with the question of what kind of nondeterminacy one faces. Third, what kinds of practical implications nondeterminacy can have is an important issue that has received surprisingly little attention in the research on hard cases. Classifying hard cases based on their practical implications rather than on metaphysical explanations enables progress in this area by shifting the focus from metaphysical explanations that have no clear practical implications to the question of what reasonable responses to nondeterminacy can be. Fourth, focusing on what a reasonable response requires instead of on the metaphysical underpinnings of nondeterminacy better aligns the classification of hard cases with what decisionmaking agents have access to. Decision making agents have no access at all to the metaphysical realm in which the true answer to the question of whether something is due to vagueness or non-conventional comparative relations resides. By contrast, decision making agents have at least more access to the realm in which the answer to the question of what a reasonable response is resides. A classification based on practical implications is thus itself more practical than a classification based on metaphysical explanations. There is, in other words, good reasons to shift the focus from metaphysical explanations to classifications based on practical implications and this is so regardless of whether our particular suggested classification is appealing or not.

Classification based on practical implications can take many directions. It seems obvious, however, to start by focusing on how it is reasonable for an agent to respond to the fact that two alternatives are nondeterminate in their ranking. How would a perfectly reasonable individual who faces a choice between two alternatives that are nondeterminate in their ranking act? The answer to this question will depend on what kind of nondeterminacy she faces and a classification of nondeterminacy should reflect this fact.

The following distinction seems to do well in capturing this:

- *Dispositive nondeterminacy*: A hard case is an instance of dispositive nondeterminacy if and only if there are two alternatives, A and B, that are both maximal and nondeterminate in their ranking with respect to F, and if and only if it is reasonable to choose any alternative that is maximal with respect to F in virtue only of the fact that it is maximal with respect to F.
- Non-dispositive nondeterminacy: A hard case is an instance of non-dispositive nondeterminacy if and only if there are two alternatives, A and B, that are both maximal nondeterminate in their ranking with respect to F, and if and only if it is reasonable to choose an alternative that is maximal with respect to F if and only if φ , where φ is a further reason in favor of an alternative that is independent of the reasons that established that A and B are nondeterminate in their ranking with respect to F.

This classification is based on the fact that it is in some cases of nondeterminacy reasonable to merely pick any maximal alternative while this is not true in other cases. In some hard cases, it is not warranted to be indifferent – i.e. what is a reasonable choice is underdetermined. For instance, it is not reasonable to be indifferent to what career to pursue if a career as a lawyer and a career as an academic are nondeterminate in their ranking. We call such cases non-dispositive. On the other hand, in some hard cases, even if F fails to fully determine a conventional comparative relation between two alternatives, A and B, it is fully determined that one can reasonably choose either. For instance, it is perfectly reasonable to be indifferent to whether one has an apple or a pear if those are nondeterminate in their ranking with respect to what afternoon snack to have. We call such cases of nondeterminacy dispositive. To make the point in a different way: Amartya Sen's proposal that it is sufficient for rational choice that an option is maximal is valid for some but not all cases of nondeterminacy and this fact can constitute a fruitful basis for a classification of comparability problems (see Sen 1997). The distinction between dispositive and non-dispositive nondeterminacy distinguishes these cases from each other.

Rather than focusing on the underlying explanation of hard cases, the distinction in terms of dispositive and non-dispositive nondeterminacy relies on the practical problems hard cases actualize (or lack thereof). Sometimes, reasons that apply to a choice underdetermine what an agent has most reason to do while it is still determined that an agent can reasonably choose a maximal alternative (dispositive nondeterminacy). This tells us something about the nature of practical reasons and how they sometimes relate to our conceptions of reasonable behavior: practical reasons do not always have to fully determine a best alternative in order to sufficiently guide reasonable behavior. However, sometimes it is the case that the reasons that apply to a choice underdetermine what an agent has most reasons to do and it is also underdetermined what a reasonable course of action is (non-dispositive nondeterminacy). This also tells us something about the nature of practical reasons and how they sometimes relate to our conceptions of reasonable behavior: in order for reasonable behavior to be at all possible when there is non-dispositive nondeterminacy, some special kind of reason that justifies determinate conventional rankings of alternatives that are nondeterminate in their ranking must exist. This seems to be a common conception, e.g., a similar point is made in less general terms by Chang when she presents what she calls 'hybrid voluntarism' according to which voluntarist reasons are reasons with this special property (Chang 2013).

It is worth pointing out that the distinction between dispositive and nondispositive nondeterminacy has analogues that are useful also for those who are committed to specific metaphysical views and prefer talking about for example indeterminacy, parity or incommensurability instead of nondeterminacy. Those who favor vagueness explanations can fruitfully draw a distinction between dispositive and non-dispositive indeterminacy to distinguish cases in which any admissible precisification is reasonable (dispositive) from cases in which one needs further reasons that justify a specific admissible precisification (non-dispositive). Likewise, those who favor non-conventional comparative relations explanations can distinguish between dispositive and non-dispositive non-conventional comparative relations to distinguish cases in which any option that is not worse than any alternative is reasonable (dispositive) from cases in which one needs further reasons to justify the selection of an option that is not worse than any alternative (non-dispositive).

3.

A classification in line with the one just sketched out raises several questions. For instance: Is it always reasonable to use conventional decision methods to discard ineligible alternatives (i.e. to discard non-maximal alternatives as unreasonable)? What sort of φ can make a choice reasonable in cases of non-dispositive nondeterminacy? Debating – or even settling – whether we should understand nondeterminacy in terms of vagueness or in terms of non-conventional comparative relations gives us no insights in these areas. By contrast, framing the phenomena in terms of dispositive and non-dispositive nondeterminacy (and other kinds of nondeterminacy) makes it easier to identify these issues and to do research on them.

Furthermore, there are purely conceptual benefits with this distinction. The upshot of distinguishing dispositive nondeterminacy from other kinds of nondeter-

minacy is that it provides a conceptual tool for delineating a kind of hard case that does not pose any serious challenges for practical reasoning and choice theory without making any direct reference to metaphysical phenomena such as vagueness or non-conventional comparative relations. In other words, some hard cases are actually *easy-to-choose-hard-to-compare-cases*.

Moreover, and as indicated already in the previous section, the distinction between dispositive and non-dispositive nondeterminacy corresponds to a distinction between how certain formal decision strategies can be brought to bear on cases of nondeterminacy. In cases of dispositive nondeterminacy there are formal decision strategies the application of which suffices to determine a reasonable choice. Decision strategies that take into account the possibility of dispositive nondeterminacy have to be defined differently depending on which explanatory framework one uses, but the choice of explanatory framework has no implication for what the decision methods imply. If one explains dispositive nondeterminacy as a special instance of when a non-conventional comparative relation obtains, one can follow Sen and replace optimality requirements with maximality requirements in rational choice theory in order to determine what is reasonable (Sen 1997). If one explains dispositive nondeterminacy as a special kind of vagueness, one can formulate a decision rule according to which all alternatives that are best on some admissible precisification are reasonable (Broome 2009; Herlitz 2019). If one explains dispositive nondeterminacy within a fitting attitudes-approach to value, one can state that it is a kind of situation in which any alternative that is highest ranked on some permissible preference ordering is reasonable (Rabinowicz 2008). Importantly, in cases of dispositive nondeterminacy maximality, best on some admissible precisification, and highest ranked on some permissible ordering suffices for a choice to be reasonable.¹⁸

By contrast, in cases of non-dispositive nondeterminacy there is no formal decision strategy the application of which suffices to determine a reasonable choice. Maximality, best on some admissible precisification and highest ranked on some permissible preference ordering are criteria that can be used also in cases of non-dispositive nondeterminacy, but instead of identifying reasonable alternatives, these criteria can at best only help one discard unreasonable alternatives when the hard case is an instance of non-dispositive nondeterminacy. The formal criteria might, in these cases, be used to reflect the side-constraints that reflect the necessary criteria for reasonable choice, but they are insufficient in that one must introduce further reasons in order to identify which of the remaining alternatives to choose.

The biggest challenge, however, classifications similar to the one sketched above

¹⁸ Things, however, become more complicated in a situation where the agent faces a sequence of choices. In such a situation the agent runs the risk of being money pumped. The decision strategies thus need to be developed in such a way that this possibility is eliminated.

will be to describe the conditions under which nondeterminacy is dispositive and under which it is non-dispositive. It should, however, be stressed that there is in this respect no difference between classifications based on practical implications and those based on metaphysical distinctions. Those who believe that some cases are hard because of vagueness owe us an answer to the question of when vagueness occurs. How are these cases separated from hard cases that are hard because we have too little information? Those who believe that some cases are hard because there are non-conventional comparative relations also owe us an answer to the question of when such relations occur? How are these cases separated from hard cases that are hard because we have too little information, or because of vagueness? That being said, it remains important to clarify the conditions under which the nondeterminacy is dispositive and those for which it is non-dispositive. A classification that is based on the practical implications of hard cases must provide a satisfactory account of these conditions. This should be taken to be an important desideratum when developing such a classification.

Although there is no space to give a definitive answer to the question 'How can a decision maker distinguish between dispositive and non-dispositive cases of nondeterminacy in practice?' here, a rough answer can be given. In light of the discussion above, certain features indicate whether a hard case is non-dispositive or dispositive. First, nondeterminacy seems to be non-dispositive when stakes are high. If A and B are maximal and nondeterminate in their ranking and if it has significant consequences whether one chooses A, B or is indifferent between them, then we have reason to believe that we are facing non-dispositive nondeterminacy. This of course leads to the question of what it means that stakes are high. It is hard to give a general answer to that question, but we believe that it is at least sometimes obvious that the stakes are high. For instance, if A and B are different career paths that an agent can embark on, the stakes are tremendously high; the choice determines a significant part of the nature of the agent's life (see Chang 1997: 23; Raz 1998: 332). Similarly, if A and B are different allocations of scarce health-resources where one alternative is better for the worse off in society and the other alternative generates significantly more health benefits overall, the stakes are very high; the choice has life and death implications (Herlitz forthcoming). Likewise in population ethics in which nondeterminacy recently has been proposed as one of the most plausible ways of avoiding the Repugnant Conclusion (see, e.g., Qizilbash 2007; Parfit 2016): if A and B are different populations of different size and where individuals have different levels of wellbeing, the stakes must be considered high; the choice concerns who will live and it will determine how good the lives of those who live will be.

Second, non-dispositive nondeterminacy seems implausible when the costs

involved in introducing new (valid) reasons are very high. One reason why it is reasonable for an agent to merely pick any maximal alternative when she, for instance, chooses an afternoon snack is that the stakes are very low, but what fundamentally makes it overly exigent to demand of her to introduce new reasons is that it is unreasonable to demand an effort of her when so little is at stake. Introducing new reasons requires at the very least identifying a valid reason, and this is too high a cost in some contexts. When the stakes are extremely small, it seems unreasonable to demand of an agent to take on any cost to make a choice. It is also worth pointing out that the costs involved in identifying a valid reason can be very high. What counts as a valid reason plausibly depends on context, and in some contexts, it is very costly to identify valid reasons that can determine which maximal alternative to choose. For instance, if the nondeterminacy arises in a social choice situation in which a large group of people will be affected by the choice, it might be suggested that the only kinds of reasons that are valid when one chooses between maximal alternatives are reasons that reflect the opinions of the affected people. Learning the opinions of a large group of people can be very costly. It seems unreasonable to demand of a social planner to take on this cost when the stakes are relatively low, for instance if the social planner decides which store to spend tax money on office supply for the public officials in.

It can now be generally suggested that nondeterminacy is dispositive if the stakes are relatively low and the costs involved in identifying a valid reason that applies to the choice and can determine which of several maximal alternatives that are nondeterminate in their ranking to choose are relatively high. Conversely, nondeterminacy is arguably non-dispositive when the stakes are relatively high and the costs involved in identifying a valid reason that applies to the choice are relatively low. Again, this is just a possible rough answer to the question of how to distinguish non-dispositive nondeterminacy from dispositive nondeterminacy. Much more can and needs to be said. For instance, what does it precisely means that the stakes are relatively low and the costs relatively high? The full characterization cannot be determined at the level of generality of this paper, but we are hopeful that more precise views can be defined in more specific areas.

4.

In this paper, we argued that the current debate around how to best explain hard cases has little bearing on how to respond to, and make reasonable choices when facing, hard cases. The debate between proponents of the Vagueness View and proponents of the Non-Conventional Comparative Relations View is little more than a disagreement about what the necessary implications of vague comparatives are. Do vague comparatives always fully determine reasonable choice or do they sometimes fail to fully determine reasonable choice? This issue has implications for whether one *calls* certain hard cases instances of non-conventional comparative relations such as parity or whether one can call all hard cases instances of indeterminacy, but it tells us nothing about how to *act* in hard cases.

In order to bring the debate on what fundamentally matters in hard cases forward – the implications for practical reasoning – we argued that a new classification that is neutral with respect to whether the Vagueness View or the Non-Conventional Comparative Relations View should be introduced. We provided a first sketch of such a classification by defining hard cases in terms of *nondeterminacy* in a way which ought to be acceptable to proponents of both of the currently dominant view. We then distinguished between two kinds of nondeterminacy. On the one hand, there is a kind of nondeterminacy which is dispositive, i.e. although some maximal alternatives are nondeterminate in their ranking it can still be fully determined what one is permitted to choose. On the other hand, there is a kind of nondeterminacy which is non-dispositive, i.e. some alternatives are nondeterminate in their ranking and it cannot be determined what one can reasonably choose.

Shifting the debate toward the issue of whether hard cases are instances of dispositive or non-dispositive nondeterminacy allows one to ask and pursue new and arguably much more pressing questions. When is nondeterminacy non-dispositive? How can one establish that an instance of nondeterminacy is non-dispositive? What is it to act reasonably in the face of non-dispositive nondeterminacy? What kinds of reasons can justifiably be brought to bear on the question of what one ought to choose when two alternatives are non-dispositively nondeterminate in their ranking? It is by answering these questions that we can make progress on how to reasonably respond to hard cases.

The main purpose of this paper is not to provide tools that decision makers can use in order to identify what sort of hard case they are facing. Instead, we wish to draw the attention to a largely overlooked distinction pertaining to the different problems hard cases might cause for choice theory and practical reasoning, and suggest that it is more fruitful to speak about hard cases by using a classification that reflects the actual problems caused by hard cases. This classification is based on the very reason why people care about hard choices, and it reflects a significant difference between different hard cases. It allows for rational choice approaches to be evaluated for specific subsets of hard cases without being evaluated at an overly general level, and it is a classification that is desirable since the alternative, currently widespread, approach is to classify hard cases in terms of conceptions of nonconventional comparative relations and vagueness that leave it entirely unspecified what the problem for practical reasoning and rational choice actually is. In this paper we have presented the structure of a new classification. Much more research needs to made in order to provide a satisfactory account. With this paper we hope that the research community can refocus from attempting to make a metaphysical distinction of hard cases to make a classification grounded on practical implications that indeed will facilitate and be relevant for decision making.

References

Andersson, Henrik & Herlitz, Anders. 2018. "A More Plausible Collapsing Principle", *Theoria*. 84, 325–336.

Andersson, Henrik. 2017. *How it all relates: Exploring the space of value comparisons*. Lund Mediatryck.

Berlin, Isaiah. 1969. "Two Concepts of Liberty." In his *Four Essays on Liberty*, Oxford: Oxford University Press.

Bognar, Greg & Iwao Hirose. 2014. *The Ethics of Health Care Rationing*. New York: Routledge.

Broome, John. 1997. "Is Incommensurability Vagueness?" In *Incommensurability, Incomparability, and Practical Reason*, ed. Ruth Chang. Cambridge, MA: Harvard University Press.

Broome, John. 2004. Weighing Lives. Oxford: Oxford University Press.

Carlson, Erik. 2010. "Parity Demystified." Theoria. 76, 119-128.

Carlson, Erik. 2013. "Vagueness, Incomparability, and the Collapsing Principle." *Ethical Theory and Moral Practice*. 16: 449-463.

Chang, Ruth. 1997. "Introduction." In *Incommensurability, Incomparability, and Practical Reason*, ed. Ruth Chang. Cambridge, MA: Harvard University Press.

Chang, Ruth. 2002. "The Possibility of Parity." Ethics. 112(4): 659-688.

Chang, Ruth. 2005. "Parity, interval value, and choice" Ethics, 115(2), 331-350.

Chang, Ruth. 2013. "Grounding practical normativity: going hybrid" *Philosophical Studies*, *164*(1), 163–187.

Chang, Ruth. 2016. "Comparativism: The Grounds of Rational Choice." In *Weighing Reasons*, eds. Errold Lord & Barry Maguire. Oxford: Oxford University Press.

Constantinescu, Cristian. 2012. "Value incomparability and indeterminacy." *Ethical Theory and Moral Practice*, *15*(1), 57–70.

Elson, Luke. 2014. "Borderline Cases and the Collapsing Principle." *Utilitas*. 26: 51–60.

Elson, Luke. 2017. "Incommensurability as Vagueness: A Burden-Shifting Argument." *Theoria*. 83: 341–363.

Fröding, Barbro & Martin Petersson. 2012. "Virtuous Choice and Parity." *Ethical Theory and Moral Practice*. 15: 71–82.

Griffin, James. 1986. *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Clarendon Press.

Gustafsson, Johan. 2011. *Preference and Choice*. Stockholm: Royal Institute of Technology.

Gustafsson, Johan. 2013. "Indeterminacy and the small-improvement argument". *Utilitas*, *25*(4), 433–445.

Herlitz, Anders. 2019. "Nondeterminacy, two-step models and justified choice." *Ethics* 129: 284–308.

Herlitz, Anders. 2020. "Nondeterminacy, cycles and rational choice," *Analysis* 80: 443–449.

Herlitz, Anders. forthcoming. "Toward a hybrid theory of how to allocate healthrelated resources", *Journal of Medicine and Philosophy*.

Hsieh, Niem-hê. 2007. "Is Incomparability a Problem for Anyone." *Economics and Philosophy.* 23: 65–80.

Hyde, Dominic. 1997. "From Heaps and Gaps to Heaps of Gluts" *Mind*, io6, pp. 641–60.

Kamm, Frances. 2007. Bioethical Prescriptions. Oxford: Oxford University Press.

Keefe, Rosanna. 2000. Theories of vagueness. Cambridge University Press.

Messerli, Michael., & Reuter, Kevin. 2017. Hard cases of comparison. *Philosophical Studies*, *174*(9), 2227–2250.

Parfit, Derek. 1984. Reasons and Persons. Oxford: Clarendon Press.

Parfit, Derek. 2016. "Can We Avoid the Repugnant Conclusion?" *Theoria*, 82: 110–127.

Qizilbash, Mozaffar. 2007. "The Mere Addition Paradox, Parity and Vagueness," *Philosophy and Phenomenological Research*, 75(1): 129–151.

Rabinowicz, Wlodek. 2008. "Value Relations." Theoria. 74, 18-49.

Rabinowicz, Wlodek. 2009. "Incommensurability and Vagueness." *Proceedings of the Aristotelian Society*, supplementary vol. LXXXIII: 71–94.

Raz, Joseph. 1986. The Morality of Freedom. Oxford: Clarendon Press.

Regan, Donald, H., (1997), "Value, Comparability, and Choice." In Chang, Ruth (ed.), *Incommensurability, Incomparability, and Practical Reason*, Harvard University Press, Cambridge, Mass., 129–150.

Sen, Amartya. 1997. "Maximization and the Act of Choice." *Econometrica*. 65: 745–779.

Sen, Amartya. 2017. "Reason and Justice: The Optimal and the Maximal." *Philosophy*, 92: 5–19.

Schoenfield, Miriam. 2014. "Decision Making in the Face of Parity." *Philosophical Perspectives*. 28: 263–277.

Wasserman, Ryan. 2004. "Indeterminacy, Ignorance and the Possibility of Parity." *Philosophical Perspectives*. 18: 391–403.

Williams, Bernard. 1981. "Conflicts of Values." In his *Moral Luck*. Cambridge: Cambridge University Press.

Williamson, Timothy. 1994. Vagueness. London: Routledge.

Krister Bykvist and Tim Campbell¹

Frick's Defence of the Procreation Asymmetry²

According to The Procreation Asymmetry: there is strong moral reason not to create a person with a miserable life but no moral reason to create a person with a happy life. Johann Frick has recently developed a novel and detailed account of this claim. The account aims to provide a theoretical explanation of the procreation asymmetry and to show that this explanation is compatible with an intuitively plausible solution to the nonidentity problem. We argue that Frick's account does not explain the procreation asymmetry but merely reaffirms it, and that his attempt to provide an intuitively plausible solution to the non-identity problem faces problems that have no clear solutions.

 $^{{}^1} Institute \ for \ Futures \ Studies, krister. by kvist@iffs.se, timothy. campbell@iffs.se.$

² We would like to thank Gustaf Arrhenius, Paul Bowman, Anders Herlitz, Karim Jebari, Julia Mosquera, and Joe Roussos for helpful comments on an earlier draft of this paper. Financial support from Riksbankens Jubileumsfond (grant M17-0372:1) is gratefully acknowledged.

1. Introduction

Suppose that if you were to create a child he would be utterly miserable. By itself, this fact seems to give you a strong moral reason not to create the child, assuming you can refrain from creating the child and that all other morally relevant considerations are equal. Almost everyone, we think, would accept this verdict.

Now suppose that you can create a child with a life well worth living, someone who would be happy and would flourish. Creating this child might have benefits for you and for others, but for the moment let us ignore such benefits. Does the fact that the child would have a life well worth living, by itself, give you a moral reason to create it? Some philosophers answer 'no'.³ They claim that you have no moral reason to create the child, or to refrain from creating it; according to these philosophers, whether you create such a child is a matter of moral indifference. Those who defend both this verdict and the previous verdict hold a position known as

The Procreation Asymmetry (PA):

(1) If a future person would foreseeably have a life that is not worth living, this in itself gives us a strong moral reason to refrain from bringing this person into existence.

(2) By contrast, there is no moral reason to create a person whose life would foreseeably be worth living, just because her life would be worth living.⁴

Some find the PA intuitively appealing. But if one wants to embrace PA, one has to deal with some important difficulties. One difficulty is to uphold the PA and at the same time have an intuitively plausible response to Derek Parfit's famous nonidentity problem.⁵ Suppose that you can decide to either create no new person, create a person with a moderately happy life or create a different person with a very happy life. All other things are equal. Intuitively, in this case, you have a moral reason not to create the person with the moderately happy life rather than the person with the very happy life. But it is difficult to accept this claim if one thinks that whether you create someone with a happy life (no matter how happy) rather than create no one at all is a matter of moral indifference. If we are morally indif-

³ Some would instead say that the answer is that the reason to create a happy person is *weaker* than the reason not to create an unhappy person (see, e.g., Holtug 2010, ch. 9). This is a weaker and arguably more plausible claim. But we will follow Frick in putting aside this weaker version of the procreation asymmetry.

⁴ Frick (2020), p. 54. Defenders of the Asymmetry include Roberts (2011), Benatar (2006), and Narveson (1976).

⁵ 1984, p. 358.

ferent between creating a very happy person and creating no one, and we are morally indifferent between creating a moderately happy person and creating no one, shouldn't we be morally indifferent between creating a very happy person and creating a moderately happy person?

Another difficulty that a PA-defender has to address is to explain why the fact that a life would be unhappy gives us reason not to create it, whereas the fact that a life would be happy does not give us reason to create it (a difficulty Frick dubs 'the Objection from Symmetry', p. 58). More specifically, the problem here is to explain this asymmetry without assuming some very similar kind of asymmetry.

In his recent paper "Conditional Reasons and the Procreation Asymmetry" Johann Frick attempts to provide a unified account of the PA that satisfactorily deals with these difficulties. Frick's main project includes three parts. The first seeks to explain the PA in terms of the notion of a bearer-regarding reason-a reason grounded in a particular well-being bearer. The second attempts to build upon this explanation by locating the PA within a more extensive normative phenomenon that includes other apparent asymmetries in reasons, such as asymmetries in the moral reasons that we have to make (or refrain from making) promises. Frick attempts to explicate this wider normative phenomenon in terms of the notion of a standard-regarding reason, of which bearer-regarding reasons are one specific instance. The third part tries to show how standard-regarding reasons can be 'captured' in terms of more familiar conditional reasons. Specifically, Frick argues, both conjuncts of the PA can be 'captured' in terms of a certain wide-scope conditional reason statement together with two inference rules. Having developed his threepart explanation of the PA, Frick attempts to show how this explanation can be combined with an intuitively plausible solution to the non-identity problem, one that avoids the implication that there is no moral reason to avoid creating a moderately happy person rather than a very happy person.

Frick's account of the PA is notable for the breadth of explanation that it purports to offer. It ties together three seemingly distinct normative notions—bearerregarding reasons, standard-regarding reasons, and conditional reasons—in an attempt to provide what is perhaps the most comprehensive explanation of the PA to date.

In this paper, we scrutinize Frick's account of the PA, as well as his proposed solution to the non-identity problem. In Section 2, we discuss the first part of Frick's explanation, which seeks to account for the PA in terms of bearer-regarding reasons. We argue that ultimately the account offered does not explain the PA but merely reaffirms it. In Section 3, we argue that the failure to explain the PA in terms of bearer-regarding reasons undermines the second part of Frick's explanation, which seeks to further explicate the PA in terms of standard-regarding reasons. In Section

4, we argue that Frick's principle of wide-scope conditional procreative reason and his two inference rules fail to explain both conjuncts the PA, and raise further philosophical problems with this account that have no clear solutions. In Section 5, we raise doubts about Frick's attempt to provide an intuitively plausible solution to the non-identity problem. Section 6 concludes.

2. Bearer-regarding reasons and the Procreation Asymmetry

The first part of Frick's explanation of the PA purports to give a unified account of both conjuncts of the PA. Most of the discussion in this first part is devoted to providing an explanation of the second conjunct—i.e. the claim that there is no moral reason to create a person whose life would foreseeably be worth living just *because* her life would be worth living.

Frick's account of the second conjunct appeals to the notion of a bearer-regarding reason—i.e. a reason that is grounded in a particular person or well-being bearer, as opposed to some state of affairs to which this person's existence might contribute positive or negative value. Frick claims that because a moral reason to confer well-being on a person is bearer-regarding, it is conditional on the existence of that person:

According to the view I advocate, our moral reasons to confer well-being on people are what I call "bearer-regarding" reasons, *and as such are conditional* on the fact that the person being benefited exists.⁶

Here the italicized text (our emphasis) suggests that a reason's being bearerregarding implies that this reason is conditional on the existence of the bearer. The idea seems to be that if no particular person (well-being bearer) exists, then there cannot be any reason "regarding" that person (well-being bearer). Since the person is what grounds the reason, it seems, the person's existence is a necessary condition of the existence of the reason.

Other passages in Frick's paper support this interpretation as well. For example, when discussing his motivation for rejecting a picture of well-being as something "to be promoted" Frick states

⁶ P. 58, emphasis added.

Replacing this picture with one according to which our reasons to confer wellbeing on people are bearer regarding, *and thus conditional* on their existence, allowed me to give a defense of the Procreation Asymmetry that strikes me as both theoretically neat and inherently plausible.⁷

Again, the italicized text (our emphasis) suggests that a reason's being bearer-regarding implies that it is conditional on the existence of the bearer. We can therefore interpret Frick as endorsing

The Bearer-Regarding View (BRV): A moral reason to confer well-being on a person is bear-regarding, i.e. it is grounded in that particular person *and is there-fore conditional* on the existence of that person.

BRV is controversial, but if true it would seem to explain the second conjunct of the Procreation Asymmetry. For BRV seems to imply that when your choice of act would determine whether some person exists, the mere fact that *if* this person were to exist she would have a life worth living does not give you a moral reason to bring her into existence. According to BRV, the existence of a moral reason to confer personal well-being is conditional on the existence of the person (or persons) on which well-being would be conferred. If you refrain from creating a person with positive well-being, i.e. with a life that is worth living, then the condition on which the existence of such a moral reason depends is not met. Thus, Frick claims that 'in a world where S herself is absent, there is no moral reason to lament the absence of S's potential happiness.'⁸

And on the same page Frick writes

Nor is there any moral reason to be exercised by the nonexistence of a potential person whose life would have been well worth living, since there is no person for whose sake we have reason to be exercised. This is the deep truth behind Jonathan Bennett's remark that while we have reason to deplore a situation where a person lacks happiness, there is no reason to deplore a situation where happiness lacks a person.⁹

But what about the first conjunct of the Procreation Asymmetry – i.e. the claim that if a future person would foreseeably have a life that is not worth living, this in itself

⁷ P. 82, emphasis added.

⁸ Frick, p. 68.

⁹ Frick, p. 68.

gives us a strong moral reason to refrain from bringing this person into existence? On Frick's interpretation of this conjunct, our moral reason to avoid creating a person with a life not worth living is *unconditional*—in other words the existence of the moral reason to avoid creating such a person is *not* conditional on the existence of that person.¹⁰ For example, as we saw earlier, Frick thinks that I have an unconditional moral reason to refrain from creating a miserable person—someone with negative well-being. This claim is very plausible. Intuitively, if I choose to create no one rather than create a miserable person, then I have *complied* with my moral reason not to create a miserable person. My reason to avoid creating a miserable person exists even if the miserable person does not exist.

But one apparent problem here is that if my reason to avoid creating a miserable person is unconditional, then this reason cannot be *bearer-regarding*, at least if we accept the characterization of a bearer-regarding reason as a reason that is grounded in a particular well-being bearer *and therefore conditional* on the existence of this bearer. If we accept this characterization, then it seems we should say that just as "there is no moral reason to lament the absence of S's potential happiness" similarly there is no reason to deplore a situation where happiness lacks a person", similarly there is no reason to favour a situation in which misery lacks a person. How, on Frick's view, can we make sense of the first conjunct of the Procreation Asymmetry?

Frick addresses this issue in a passage on page 69 that continues in an accompanying endnote. Here is the passage:

[N]othing I have said in this section challenges the first conjunct of the Procreation Asymmetry, i.e. the claim that we *do* have moral reason to avoid creating lives that will foreseeably be miserable. For the world in which I create S with a miserable life is precisely a world in which ... this fact *is* of moral concern, *since S exists at that world*.

Here Frick claims that *if I create* S with a miserable life, then this fact is of moral concern. But that is not sufficient to establish that I have an *unconditional* moral reason to refrain from creating S with a miserable life—i.e. a moral reason whose existence does not depend on S's existence. However, in the endnote to the passage, Frick continues

¹⁰ For example, he considers and rejects a potential explanation of the Procreation Asymmetry on the grounds that this explanation "lacks the resources to explain how I could have an unconditional reason *not* to create a child whose life will be irredeemably miserable" (p. 73).

Moreover, it is consistent with the view I am defending that I have moral reason to avoid a state of affairs in which S exists and is miserable even if S does not currently exist (and will never exist, if I act rightly). ... According to my view

(1) I can have S-regarding reasons *against* bringing about a world *w* only if S exists in *w*.

(...) But (1) is compatible with

(2) I can have S-regarding reasons *at* a world *w1* against bringing about a world *w2*, even if S does not exist in *w1* but only in *w2*.

However, there are at least two problems here. First, neither in the main passage on pages 69 nor in the accompanying endnote does Frick provide an *explanation* of the first conjunct of the Procreation Asymmetry. That is, he does not provide an account of our unconditional moral reasons to avoid creating miserable people. As we saw, in the passage on pages 69, Frick appeals to an example that cannot explain why we have such unconditional reasons. And in the endnote to the passage, Frick seems to claim only that his account of the second conjunct of the Procreation Asymmetry *is compatible with* claim (2)—i.e. the claim that I can have S-regarding reasons *at* a world *w1* against bringing about a world *w2*, even if S does not exist in *w1* but only in *w2*.

Second, and more importantly, claim (2) implies that we can have bearer-regarding reasons that are unconditional. For example, claim (2) implies that even if it is not the case that person S exists (now or in the future), I can still have an S-regarding reason against creating S with a miserable life. Assuming that an S-regarding reason is a *bearer-regarding* reason, this implies that I can have a bearer-regarding reason that is not conditional on the existence of the bearer. It is in fact crucial for Frick that he acknowledges such unconditional bearer-regarding reasons, for otherwise he will fall afoul of actualism, which he explicitly rejects.¹¹ Actualism claims that we should only be concerned with actual people, the people who will ever exist. But since who the actual people are will often depend on what we do, as in the case where we can create a miserable person, we get the result that whether we have reason to do an action can depend on whether we will do it. In particular, if you were to create S who would have a miserable life, you would have S-regarding reason not to do so, but if you were not to create her, you would not have any S-regarding reason not to do so. This normative variance is problematic, since it makes it difficult to deliberate and can also lead to a certain kind of dilemma, which Frick acknowledges.¹²

¹¹ Frick, 61-62.

¹² Ibid. These problems are more thorougly discussed in Bykvist (2007).

Now, if it is true that we have an unconditional bearer-regarding reason not to create a person with a miserable life, then Frick owes us a further explanation as to why I cannot have an unconditional bearer-regarding reason to create a person with a life *worth living*. The explanation that he initially offers is that my reason to confer well-being on a person is bearer-regarding and *as such* is conditional on the existence of the person. But if some bearer-regarding reasons are unconditional, then it is not the case that bearer-regarding reasons *as such* are conditional on the existence of the bearers that they regard. So we still need an explanation as to why *only some* of our bearer-regarding reasons, namely those we have to confer well-being on people, are conditional on the existence of those people. Since Frick offers no such explanation, it seems he is merely *reaffirming* the Procreation Asymmetry, not *explaining* it.

3. Bearer-regarding reasons and standard-regarding reasons

We have argued that Frick's account of the Procreation Asymmetry in terms of bearer-regarding reasons just reaffirms the procreation asymmetry. In this section, we argue that this fact undermines Frick's theoretical explanation of the Procreation Asymmetry in terms of standard-regarding reasons. Recall that in addition to providing an independent account of the Procreation Asymmetry in terms of bearer-regarding reasons, Frick seeks to further explain the asymmetry by locating it within a wider normative phenomenon that includes apparent asymmetries in moral reasons related to promise-keeping. He does this by appealing to the notion of a *normative standard*. Frick writes:

Characterizing our moral reasons with regard to procreation in terms of this notion gives us a simple way of capturing both conjuncts of the Asymmetry Intuition. The concept of normative standard also enables us to render explicit the structural parallels with other normative asymmetries, such as the promissory asymmetry we considered above. ... A normative standard ... is a criterion of evaluation that applies to those outcomes of an agent's actions that fall within the *scope* of the standard. An outcome can either *fail* to satisfy the standard, in which case there are "standard-regarding" reasons to avoid this outcome, or it can *pass* (or *satisfy*) the standard, in which case there are no standard-regarding reasons against bringing it about.¹³

¹³ Frick, pp. 69–70.

In the case of promising, for example, there is a promisory standard that applies only to those outcomes in which one makes a promise. Although one has reason to keep one's promise given that one has made (or will make) a promise, Frick claims that one has no moral reason (issuing from the promisory standard) to merely increase the number of kept promises by making a promise and keeping it. As we saw, Frick believes that a similar claim is true of creating well-off people. Although I have a moral reason to confer well-being on a person *given that I create this person,* I have no moral reason to create a happy person just because there would then exist an additional happy person. In general, Frick thinks, I do not have any standard-regarding reason to satisfy a normative standard merely by bringing about an outcome in which the relevant standard applies and is satisfied.

On the other hand, Frick thinks that I have an unconditional standard-regarding reason to avoid bringing about an outcome in which a normative standard applies and I fail to comply with the standard:

Suppose ... that I can foresee that, having made you the promise, I will be unable to keep it. Knowing this, I have a standard-regarding reason to *avoid making* the promise in the first place. More generally, there is a standard-regarding reason to avoid bringing about an outcome to which a normative standard applies and in which I am unable to satisfy that standard.¹⁴

This passage suggests that I have a standard-regarding reason to avoid bringing about an outcome in which I fail the relevant standard even if *in fact* I do not bring about an outcome in which the standard applies. If this is correct, then I have an unconditional standard-regarding reason to avoid failing a normative standard, even though I have no standard-regarding reason whatsoever to ensure that a normative standard applies in the first place and is satisfied. Since standard-regarding reasons have this asymmetrical structure, it may seem that we can explain the Procreation Asymmetry in terms of standard-regarding reasons, where the relevant normative standards are procreative ones.

However, Frick (correctly, we think) points out that because standard-regarding reasons have this asymmetrical structure,

it would have been question-begging to appeal directly to the notion of normative standard in seeking to explain the Procreation Asymmetry, without first giving an *independent* argument for the claim that we have no moral reasons to create a person just because her life will be worth living. That is why the argument of the

¹⁴ Frick, p. 70–71.

preceding section, in defence of the second conjunct of the Procreation Asymmetry, was indispensable. With that argument in place, I could *then* appeal to the notion of normative standard to provide a *unified* account of both conjuncts of the Procreation Asymmetry, without begging any questions. (72–73)

But if, as we have argued, the independent argument in terms of bearer-regarding reasons just reaffirms the Procreation Asymmetry, then it seems that Frick's further explanation of the Procreation Asymmetry in terms of the notion of normative standard is still question-begging. To show that the Procreation Asymmetry can be explained in terms of the notion of normative standard *without* begging the question, Frick would need to provide an explanation for his claim that although we have unconditional bearer-regarding reasons to avoid creating people with lives not worth living we have no unconditional bearer-regarding reasons to create people with lives worth living. But as we saw, he does not provide such an explanation.

4. Standard-regarding reasons as wide-scope conditional reasons

Frick claims that standard-regarding reasons can be 'captured' in terms of the more familiar *conditional* reasons (p. 73). More specifically, he thinks that standard-regarding reason to procreate can be captured by the following wide-scope conditional reason statement:

The Threshold Requirement (TR) I have a moral reason to (if I create a new person, make it the case that this person's life is worth living)

Moreover, he argues that TR in conjunction with two inference rules accounts for the asymmetry by providing 'an explanation of both conjuncts of the Procreative Asymmetry' (p. 75). The inference rules are these:

Rule 1 I have reason to (if I do p, do q). It is unavoidable that I do p.

I have reason to do q.

Rule 2 I have reason to (if I do p, do q). It is unavoidable that I do not do q.

I have reason not to do p.

He explains the first conjunct of the PA by showing that TR and Rule 2 together entail that we have reason not to create a person with a miserable life. The argument is straightforward. Suppose that it is unavoidable that I will fail to make a certain potential person's life worthwhile, because if I create her, she will unavoidably have a miserable life (and if I do not create her, she will fail to have any life at all). Then TR and Rule 2 together entail that I now have reason not to create the person. This reason holds unconditionally, no matter whether I in fact will create the person or not. So, we have accounted for the first conjunct of the PA.

However, TR and Rule 2 commit you to some problematic verdicts that the PA does not. Suppose that it is unavoidable that I will fail to make a certain potential person's life worthwhile, not because she will unavoidably have a miserable life but because she will unavoidably have a *neutral* life if I create her. Then TR and Rule 2 together again entail that I now have reason not to create the person. But this seems questionable. After all, what tells against creating a neutral life, if all other things are equal? This problem can be fixed by revising TR thus:

I have a moral reason to (if I create a new person, make it the case that this person's life is *at least neutral*)

What is more problematic is how Frick accounts for the second conjunct. His argument is this:

(..) there is *no* unconditional reason under the Threshold Requirement to create a child, just because I will be able to give it a life worth living. Rather, any reason to confer well-being on the child is conditional on that child's existence. Only once the proposition "I create a child" is unavoidably true — typically, once I have created the child — does Rule 1 allow me to detach an unconditional moral reason to make it the case that the child has a life that is worth living. (75)

As stated, this argument is flawed. Recall that the second conjunct of the PA says that there is no reason to create a happy person. It is true that TR and Rule 1 do not entail that there is a reason to create a happy person. But nor do they entail that

there is *no* reason to create such a person. We have to distinguish not entailing that there is a reason to A and entailing that there is no reason to A. What Frick wants is the latter, but he only gets the former. Similarly, TR and Rule 1 do not entail that *any* reason to confer wellbeing on a person is conditional on that person's existence.

The problem is that TR only states *a* wellbeing-regarding reason to do something in cases of procreation; it does not say that it is the *only* such reason.¹⁵ The problem cannot be fixed just by adding the claim that TR provides the only wellbeing-regarding reason applicable in cases of procreation, because Frick thinks that other such reasons are applicable in these cases (see below).

There is a further issue, namely, that TR itself says that you have done something you have reason to do if you create a happy person. If you create a happy person, you thereby make true the conditional 'if you create her, you make her life at least neutral'. So, you have thereby done something you have reason to do, viz, if you create her, make her life at least neutral, and this reason is unconditional. In short, you comply with the TR-reason if you create a happy life.

Of course, you have also done something you have reason to do if you *fail* to create a happy life. So, no matter whether you create the person, you do something you have reason to do - you comply with the TR-reason. This TR-reason has to be strong, for the first conjunct of the of the asymmetry says that we have a *strong* reason not to create a life that is not worth living (54). If the TR-reason is supposed to provide this reason, we have a case where no matter what you were to do, you would do something you have strong positive reason to do. But the intuition behind the asymmetry is not that procreation is a 'delightful dilemma', like a choice between two equally blissful futures (Bedke 2009, 681). The intuition was rather that if you were to create a happy person, you would do something you have *no* positive reason to do.

Frick could say that the reason we comply with when we create a happy life is not grounded in the wellbeing of the person, and so it does nothing to threaten the asymmetry, which in its official formulation says that we have reason not to create unhappy lives *because* of their unhappiness, but no reason to create happy lives *just because* of their happiness (54). But this cuts both ways. Since the TR-reason itself says nothing about what grounds the reason, it does not itself say that we have reason not to create an unhappy life *because* of the unhappiness. This is another reason to be sceptical that TR (plus the inference rules) account for the asymmetry, as it is formulated.

¹⁵ It is true that it is called a *requirement*, but the requirement has to do with the threshold that pertains to the *content* of the reason, the threshold between lives worth living and lives not worth living.

A better option for Frick is to say that the wide scope reason is a *contrastive* reason; TR should be replaced by this principle:

I have a moral reason to (if I create a new person, make it the case that this person's life is at least neutral) *rather than* not (if I create a new person, make it the case that this person's life is at least neutral).

This principle avoids the problem with the delightful dilemma, since it is not applicable in two-options cases where I either create a happy person or do not create her. To have a contrastive reason to do A rather than B you need to have both A and B as options. But in the case where I can either create a happy person or not do so, it is not an option for me to *not (if I create a new person, make it the case that this person's life is at least neutral)*, i.e., to create a new person and not make it the case that her life is at least neutral. On the other hand, the principle is still applicable in two-options cases where I either create an unhappy person or do not create her, since I can make true (if I create a new person, make it the case that this person's life is at least neutral) by simply not creating her.

Even if going contrastive will solve the problem with the delightful dilemma, Frick is still saddled with the problem of explaining the second conjunct of the Asymmetry. Since the contrastive version of TR is not applicable to cases where we can either create a happy person or not create her, it does not entail that there is no contrastive reason to create a happy person in these cases. But in order to explain the second conjunct of the PA, we need it to entail that there is no such reason.

5. The non-identity case and the Principle of Standard Selection

Frick argues that in order to deal with cases in which you can choose between creating a happy person and creating the same person with an even happier life, we need to assume the *principle of greater satisfaction*.¹⁶ This principle says, roughly, that you have a contrastive reason to bring a person into a happy existence rather than a less happy existence, if you are going to create the person at all. But he points out that this principle together with TR, Rule 1, and Rule 2 does not help us in nonidentity cases where we can choose between creating nobody (Nobody), creating a person with a moderately happy life (Good), and creating a *different* person with a very happy life (Great). In order to deal with these cases, Frick introduces yet

¹⁶ Frick, p. 77.

another principle, *the principle of standard selection*, that roughly says that you have a contrastive reason to bring about outcome O1 rather than outcome O2, if you have a choice between bringing about O1 to which a standard X applies and bringing about outcome O2 to which a different standard Y of the same kind applies, and the degree to which O1 satisfies X is greater than the degree to which O2 satisfies Y (and all else is equal). Applied to non-identity cases, the principle says that in a choice between creating a person who will have a moderately happy life and creating a person who will have a very happy life, we have a contrastive reason to do the latter, since the welfare standard for the very happy person is satisfied to a greater degree than that of the moderately happy person.

Frick thinks this principle's application to procreation cases can be justified if we adopt the following principle, which he thinks 'meshes' well with his general ideas about conditional reasons:

(..) conditional on creating any new person, we have moral reason to want her life to go as well as possible in an absolute sense, unconstrained by questions of practical feasibility.¹⁷

Frick claims that this means that we have reason to *regret* having created a person who is moderately happy because of a serious congenital ailment. It also means that if you create a very happy person in the non-identity case, you 'ensure that the world is *closer* to the world as you have reason to want it to be, conditional on having the child, than would be the case, if you create the less happy child'.¹⁸ It is this fact, Frick claims, that gives you a reason to choose Great over Good.

One obvious problem with this principle is that closeness is not defined if a person's happiness is not bounded: there is no level of happiness that the person cannot exceed. This seems plausible if we are talking about *metaphysical* possibility. So, in order to avoid this problem, Frick has to adopt some notion of possibility that comes with a cap for maximum happiness for the person without being the same as practical feasibility. But exactly how far from actuality are we supposed to go in modal space when defining this notion of possibility (technological possibility, psychological possibility, human possibility, or what)?

Another issue here is that Frick introduces a new concept, that of a conditional reason to *care*, which is supposed to be absolute in the sense of caring about how close the wellbeing of a person comes to some ideal level. He also makes explicit, which he did not do before, that this reason to care provides reason to act. Unfortu-

¹⁷ Frick, p. 80.

¹⁸ Frick, p. 81.

nately, he does not say what he means by the conditionality in reason to care. Previously, when he discussed reason to act (i.e., to create people), he understood the conditionality of this reason in terms of wide-scope reasons with conditional content. Is this how we should understand the conditionality in the reason to care as well? If we do, then the reason to care should be formulated as

You have reason to (if you create a person, you want her life to go as well as possible)

But this would imply, by Rule 2, that you have reason not to create S if you unavoidably will not want her life to go as well as possible, perhaps because you will be dead later or suffer brain damage (independently of creating S), even though S would lead a fantastic life.

An alternative wide-scope formulation would be this:

You have reason to want that (if you create a person, her life goes as well as possible)

But this would imply, by an inference rule analogous to and as plausible as Rule 2, that you have reason to want that S is not created if it is unavoidable that she will lead a life that is not maximally good for her in absolute terms, even though her life would be fantastic and not contain any bad patches. The inference rule is this:

You have reason to want that (if p, then q).

It is unavoidable that not-q.

You have reason to want that not-p.

What is even worse, you would have *overall* reason to want that S is not created, since Frick would not want to postulate any S-regarding reasons *against* wanting not to create a happy person (and we can assume that all other things are equal). If a reason for you to want A, gives you a reason to do A, which seems to be in line with what Frick says, we also get the unappealing conclusion that we have reason *not* to create the person, despite the fact she would have a uniformly splendid life. Again, this reason would be overall, since Frick does not think there are any S-regarding reasons against not creating a happy person (and we can assume that all other things are equal).

These problems suggest that it is best to go for a narrow-scope understanding of the conditional reason to care:

If you create a person, you have reason to (want that her life goes as well as possible).

(where this is read as a counterfactual conditional, so it can be applied non-vacuously to cases where you in fact will not create a person)

This formulation does not have the problems just discussed. The step from these narrow-scope reasons to the contrastive reason to choose Great over Good would then be something like this. Suppose that you can either create S or T. If you were to create S, she would be happy but not maximally so. So, you would then have reason to feel regret about that, for S's sake. If you were to create T, she would be very happy but not maximally happy. So you would then have reason to feel regret about that, for T's sake. But the regret you would have reason to feel is weaker, since T would be better off than S and thus 'closer' to being maximally happy. In a choice between doing something that you would have a reason to feel regret about and creating something that you would have a reason to feel regret about and creating something that you would have a reason to feel regret about, you have a contrastive reason to choose the latter rather than the former.

One problem here is that it is implausible to say that you would have any reason to feel regret if you created a person with a fantastic life that would be uniformly good just because it would not be maximally happy.

Furthermore, this account seems to entail that you have a reason *not* to create a very happy person. If you create the person and make her very happy but not maximally so, you will have a reason to regret your choice, for her sake, whereas if you do not create her there is no reason to regret your choice, for her sake, since she does not exist. Surely, in a choice between something that gives you reason to regret your choice and something that does not, you have a contrastive reason to do the former rather than the latter.

Could the problem be fixed by talking about reason to *favour* the situation of happy people instead? Then the story would go something like this. Suppose that you can either create S or T. If you created S, she would be happy. So, you would have reason to favour this for S's sake. If you created T, she would be very happy. So you would have reason to favour that, for T's sake. But the favouring you would have reason to feel is stronger, since T would be better off than S and thus 'closer' to being maximally happy. In a choice between doing something that creates something you would have a reason to favour and creating something that you would have reason to favour even more, you have a contrastive reason to choose the latter rather than the former.

But then it is difficult to explain why in a choice between creating a happy life and not creating the life you do not have *any* reason to do the former over the latter. After all, if you create the person, you will have reason to favour her life. But wouldn't this generate a reason to create the person rather than not create her, especially if we assume that you have a reason not to create an unhappy person over doing so because if you were to create her, you would have reason to disfavour her situation?

Of course, we could just state that there is an asymmetry here: the fact that you would have reason to disfavour having created an unhappy person gives you a contrastive reason not to create the person rather than to create her, but the fact you would have reason to favour having created a happy person does not give you any contrastive reason to create the person rather than not create her. But then we are back to just stating the asymmetry and not accounting for it. Furthermore, there is no longer any need to adopt wide-scope reasons to act. We can just state the theory in terms of narrow-scope reasons to favour or disfavour and some bridge principles, linking reasons to care to contrastive reasons to do.

6. Conclusion

Frick's account of the Procreation Asymmetry is the most comprehensive to date. It purports to give a unified non-question-begging explanation of the PA in terms of three different interrelated normative notions—bearer regarding reasons, standard-regarding reasons, and conditional reasons—an explanation that does not preclude an intuitively plausible response to the non-identity problem.

We have scrutinized each part of Frick's account of the PA as well as his proposed solution to the non-identity problem. Frick's appeal to bearer-regarding reasons does not explain the PA, but merely reaffirms it. Specifically, it fails to explain why, if one has an unconditional bearer-regarding reason not to create a miserable person, one cannot similarly have an unconditional bearer-regarding reason to create a happy person. The failure to provide such an explanation undermines Frick's attempt to further explain the PA in terms of standard-regarding reasons. Frick himself acknowledges that without an independent explanation of the PA in terms of bearer-regarding reasons, the explanation in terms of standard-regarding reasons is question-begging. Moreover, Frick's principle of wide-scope conditional procreative reason and his two inference rules fail to explain both conjuncts the PA, and raise further philosophical problems with his account that have no clear solutions. Finally, we have raised doubts about Frick's attempt to provide an intuitively plausible solution to the non-identity problem. Specifically, we argued, Frick's introduction of a conditional reason to *care*, which is supposed to be absolute

in the sense of caring about how close the wellbeing of a person comes to some ideal state, raises more problems than it solves, or reaffirms the asymmetry and makes the wide-scope conditional reasons to act otiose.

References

Arrhenius, Gustaf. 2003. "The Person-Affecting Restriction, Comparativism, and the Moral Status of Potential People". *Ethical Perspectives* 10(3-4): 185–95.

Bedke, Matthew S. 'The Iffiest Oughts: A Guise of Reason Account of End-Given Conditionals'. *Ethics* 119, pp. 672–698.

Benatar, David. 2006. *Better Never to Have Been: The Harm of Coming into Existence*, Oxford: Clarendon Press.

Bykvist, K. 'Violations of normative invariance: Some thoughts on shifty oughts', *Theoria*, Vol. 73, Issue 2, June, 2007, pp. 98–120.

Frick, Johann. 2020. 'Conditional Reasons and the Procreation Asymmetry'. *Philosophical Perspectives* 34(1): 53–97.

Holtug, Nils. Persons, *Interests, and Justice*. New York: Oxford University Press, 2010.

Narveson, Jan. 1976. "Moral Problems of Population". In Bayles, ed., *Ethics and Population*, Cambridge, 59–80, MA: Schenkman.

Parfit, Derek. 1984. Reasons and Persons. Oxford: Oxford University Press.

Roberts, Melinda. 2011. "The Asymmetry: A Solution," Theoria 77: 333-367.

Studies on climate ethics and future generations, vol. 1 Working paper series 2019:1–11

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

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

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