How to neglect the long term

Hayden Wilkinson (Global Priorities Institute, University of Oxford)

Global Priorities Institute | September 2023

GPI Working Paper No . 11-2023

How to neglect the long term[*]

Hayden Wilkinson

Last updated: August, 2023

Comments welcome: hayden.wilkinson@philosophy.ox.ac.uk

Abstract

Consider *longtermism*: the view that, at least in some of the most important decisions facing agents today, which options are morally best is determined by which are best for the long-term future. Various critics have argued that longtermism is false—indeed, that it is *obviously* false, and that we can reject it on normative grounds without close consideration of certain descriptive facts. In effect, it is argued, longtermism would be false *even if* real-world agents had promising means of benefiting vast numbers of future people. In this paper, I develop a series of troubling impossibility results for those who wish to reject longtermism so robustly. It turns out that, to do so, we must incur severe theoretical costs. I suspect that these costs are greater than simply accepting longtermism. If so, the more promising route to denying longtermism would be by appeal to descriptive facts.

Keywords: *longtermism; depletion; person-affecting views; aggregation; separability; anonymity; extinction risk.*

[*] For helpful comments on various drafts of this paper, I thank Tomi Francis, Hilary Greaves, Jakob Lohmar, Rhys Southan, and Timothy L. Williamson.

1 Introduction

Our future, in its entirety, may be far bigger than our present. As of the beginning of 2023, approximately 8 billion humans are going about their lives. Whereas, over the centuries and millennia ahead of us, vastly more human lives may be lived.[1] As long as we do not wipe ourselves out too soon—and even relative pessimists suggest that it is more likely than not that we survive (see, e.g., Ord, 2020, p. 167)—our descendants may well outnumber us by more than 100,000 to 1.[2]

This suggests that the moral stakes of decisions concerning the future are high. This need not be because it is important to *ensure* that a long future for humanity comes about in the first place, nor need it be morally important to make more people exist. No, the mere fact that there are likely to be vast numbers of future people, regardless of what we do, is enough to raise the stakes astronomically. After all, no life counts for more or less than another merely due to superficial characteristics like the person's race, sex, nationality, or the circumstances of their birth. Merely whether someone is born in the year 1960 or 2960 does not change the moral value of their well-being, nor does it change the importance of aiding them *if* we can do so with equal ease and predictability.[3] So, if many more people are born in the period 2100-100,000 CE than in the period 2000-2100 CE, the stakes are potentially far higher when influencing the well-being of everyone in the first group than when only influencing the latter group.

This is one (partial) motivation for a view known as *longtermism*: roughly that, at least in some of the most important decisions facing agents today, which options are morally best is determined by which are best for the long-term future (from Greaves and MacAskill, 2021, p. 3). At least as I am interested in it, this is an *axiological* claim: one about which outcomes, or risky prospects, are *better* than others. It is not a *deontic* claim: it doesn't state that we *ought* to do what is best for the long-term future, and certainly not that we ought to do so in *every* decision we face, nor to do so by whatever means are necessary.[4] And, crucially, it may well be a *contingent* claim: since it refers to the decisions we actually face and the options we actually have, it may well depend on what the world is *actually* like.

[1] An analogous claim holds for the even more numerous non-human animals with whom we share a planet—very many of them exist right now, but many times more will exist over the course of the future. Where I talk of 'people' throughout this paper, this may be interpreted either as including only humans or as including those within a much larger class of moral patients.

[2] Wolf and Toon (2015) give us one billion years until the Earth becomes uninhabitable—3,000 times longer than *homo sapiens* has existed to date. For arguments that this does *not* tell us that the expected number of future people is very large, see Thorstad (2023a,b). For what I take to be a persuasive counter-argument, see Chappell (2023).

[3] Such moral impartiality to such features position in time is defended by Sidgwick (1907, p. 414), Ramsey (1928, p. 541), Parfit (1984, §121 & Appendix F), and Cowen and Parfit (1992), among others.

[4] Indeed, even in the deontic version of longtermism offered by Greaves and MacAskill (2021, pp. 26-9), we are only required to do what is best for the long-term future in "...the most important decisions facing agents today, [in which] the axiological stakes are very high, there are no serious side-constraints, and the personal prerogatives are comparatively minor." In such decisions, which might even be limited to just those in which we decide how to allocate our charitable donations and other altruistic efforts, there is no clash with common non-consequentialist deontic principles.

Is longtermism true? Various philosophers and other commentators have argued that, no, it isn't. Indeed, many have suggested that denying it is a straightforward matter—that it would *only* hold on supposedly implausible normative views. Simply deny consequentialism, utilitarianism, a total view of moral betterness, or expected value theory, and longtermism is clearly false.[5] Or so it is suggested (e.g., by Cremer and Kemp, 2021; Setiya, 2022; Stock, 2022; Lenman, 2022; Henning, 2022; Wolfendale, 2022; Walen, 2022; Crary, 2023; Adams et al., 2023; Plant, 2023; Bramble, 2023).[6] Indeed, each of the cited authors claims that the falsity of longtermism is no close thing—that it would still be false if our descriptive circumstances were slightly different, if the number of future people were greater, or if our opportunities to aid them were more promising. On primarily normative grounds, it is suggested, we can *robustly* deny longtermism—we can deny it without closely examining the descriptive facts, because it would remain false even if those facts were different.

The aim of this paper is to establish just how tenable a position this is. I am interested in just what it takes to deny longtermism out of hand, without leaving the philosophical armchair to examine descriptive facts that are not obvious to us. Establishing this is of interest for two reasons. The first is that recent arguments against longtermism (as cited above) may prove too much; they may commit us to unexpected, unsavoury implications, and so should be rejected. The second reason is that it would be useful to know whether longtermism is indeed true. Although this question won't be answered directly by establishing how difficult it is to robustly deny longtermism, we may gain insight into where we should look: if longtermism cannot be robustly denied at the level of moral theory (at least not without unsavoury implications) then, to determine whether it is true, we need to look to the descriptive facts.

The discussion is structured as follows. In the next section, I present a principle that we must deny to deny longtermism robustly, namely *In-Principle Longtermism*. There follows a brief tour of the formalism to be used throughout. In Sections 4 and 5, I present a series of general impossibility results for those who wish to deny In-Principle Longtermism; deny it and one must accept one or another troubling implication.

Section 6 constitutes a slight detour, but hopefully an instructive one. In practice, many self-avowed longtermists advocate not only for improving the future, but also for *particular* interventions that they claim would improve the future (*ex ante*). MacAskill (2022, ch. 8), for instance, advocates

[5] This claim, at least, has already been shown to be false. Tarsney and Thomas (2020), for instance, make the case for longtermism based on averageism, on egalitarianism and on other non-totalist views of moral betterness; Thomas (2022a) shows how various theories of betterness upholding the well-known procreation asymmetry lead to longtermism; Buchak (2022); Pettigrew (2022) each show how risk-weighted expected utility theory can lead to longtermist conclusions; and Greaves and MacAskill (2021, §6) notes that a variety of other normative theories do too.

[6] Representative quotes include: "Longtermists ... argue that it's always better, other things equal, if another person exists, provided their life is good enough. That's why human extinction looms so large." (Setiya, 2022); "Longtermists...[maintain] that any additional person who lives makes the world better, as long as the person enjoys adequate wellbeing." (Crary, 2023); "...longtermists adopt substantial utilitarian commitments in arguing for maximizing the well-being of all." (Adams et al., 2023); and "The non-utilitarian case for strong longtermism is, for now, weak." (Cremer and Kemp, 2021).

for various actions that mitigate risks of human extinction in the near term, while also likely improving future generations' quality of life conditional on survival. Several of the results I present in earlier sections also bear on whether such actions are improvements, and seem to suggest that they are. In Section 6, I determine which results bear on extinction risks, and which might be modified slightly to do so.

As I conclude in Section 7, given the various impossibility results, it is not so easy to deny longtermism as robustly as many critics suggest—to deny it without regard to certain descriptive details. Likewise, it is not easy to deny on normative grounds the further claim that extinction risks are worth mitigating. As I show, philosophically speaking, it is not so easy to neglect the long term. To reject longtermism, I suggest, we must ultimately appeal to descriptive claims, whose truth is far from obvious.

2 In-Principle Longtermism

To deny longtermism robustly, without any appeal to (unobvious) descriptive facts, what exactly must we deny?

This depends on what descriptive facts are considered unobvious, and these cannot simply include *all* descriptive facts. If they did, longtermism would be nigh impossible to deny. (Try denying longtermism if, for some bizarre reason, we simply had no practical means of benefiting anyone *except* those in the long-term future!) I have in mind a denial of longtermism somewhat less robust than this. So, I will hold fixed some 'obvious' descriptive facts on which our denial of longtermism might still rely.

The first such obvious fact is: that in any given decision we have some option by which we can benefit present (or near future) people with high probability. Indeed, for simplicity, I will assume that we always have an option to benefit *all* present people, by just as much as we can otherwise benefit future people, and with probability 1. Although this is neither true nor obvious, this is a generous assumption—being able to benefit more present people, and with higher probability, can only make longtermism easier to avoid.

The second such obvious fact is: that any option by which we might improve the lives that are lived in the long-term future has only a low probability (of at most, say, 1 in 1,000) of doing so. [7] As for the first, this is a generous assumption. And it seems accurate of most agents' real-world attempts to greatly improve or improve the long-term future: for instance, suppose an individual

[7] Perhaps this does not hold of all options by which we might improve the long-term future. One plausible exception is to engage in *patient philanthropy*: setting aside resources (most easily, money) to be used by future decision-makers when needed most, and to accrue in value in the meantime. For detailed discussion, see Trammell (2021).

who donates a modest sum to an organisation that advocates against the proliferation of nuclear weapons due to the deleterious long-term effects of a potential nuclear war; their donation is unlikely to prevent such long-term effects, both because a modest donation rarely makes much difference to the activities of advocacy organisations, and because nuclear wars may be unlikely to occur in the first place. It seems at least plausible that the probability of their individual donation preventing a nuclear war is 1 in 1,000 or less.

The third such obvious fact that I will assume is: that, in outcomes where we do improve the long-term future, the identities of all of our future beneficiaries are altered. After all, in practice, any action with so great an impact on human history that it changes the well-being of vastly many future people must also change many of the circumstances of many people's lives, both future and present. These include the circumstances under which humans conceive children; change these even slightly and different combinations of sperm and egg will meet, and so different children will be born. So, in practice, any actions with long-lasting, widespread effects are guaranteed to be identity-affecting in this sense; resulting in an (almost) entirely different population of people alive in several centuries (and beyond).[8] If longtermism is to hold in practice, it must hold even with this phenomenon present. And to keep things simple we can assume (again, generously) that *no* future people obtain higher well-being without also having their identities changed.[9]

To deny longtermism robustly, without close regard to our descriptive circumstances, we must deny *In-Principle Longtermism.* This is the condition that longtermism *would* be true if the unobvious descriptive facts were favourable to it, even if the three 'obvious' facts listed above are not.

> *In-Principle Longtermism:* There is some pair of prospects can described by P_{Present} and P_{Future}, for some outcome O and some probability $p \leq \frac{1}{1000}$, and of which P_{Future} is better than P_{Present}.
>
>> P_{Present}: With probability 1, an outcome in which all m present people have higher well-being than in O.
>>
>> P_{Future}: With probability $1-p$, outcome O; with probability p, an outcome in which the same present people exist as in O and have the same well-being, and in which the set of future people who exist is entirely disjoint from that in O (and perhaps have higher well-being than the future people in O).

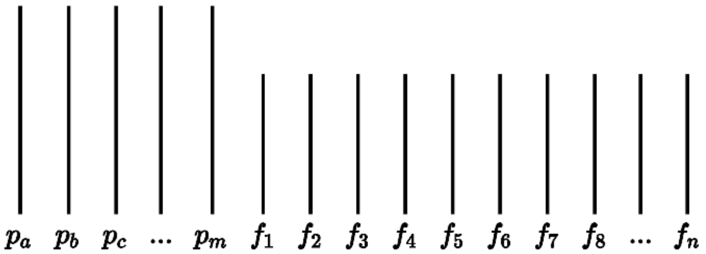
If In-Principle Longtermism is true—if there is some such P_{Future} that is better than some such

[8] This observation was made by Parfit (1984, §119&123), and has featured heavily in recent discussions of cluelessness (Greaves, 2016; MacAskill and Mogensen, 2021). A similar claim was made much earlier, although with a theological rather than biological justification, by Leibniz (2005, pp. 101-7).

[9] A further realistic assumption that I won't make just yet is that changing the well-being of many future people will inevitably also change the number of future people who exist (by much the same mechanism as it changes the identities of those future people). I'll consider the case for In-Principle Longtermism under this assumption in Section 5.

P_{Present}—then longtermism (as defined above) *could* be true. Some descriptive facts would still need to obtain for longtermism to hold—we would still need agents to actually face such pairs of options in their most important decisions, and so need the probabilities and stakes of their options to be appropriately high. But In-Principle Longtermism opens up the possibility of longtermism. And even this, I take it, will be unappealing to the critics of longtermism cited above, as each rejects it without any appeal to any unobvious descriptive facts.

To see what it takes to avoid In-Principle Longtermism, it will be helpful to also consider another, more specific claim that many critics find unappealing. This will be a claim about how to compare a particular pair of outcomes—*outcomes*, not prospects, so there is no longer any risk involved—of which one offers greater well-being to present people while the other offers greater well-being to whoever exists in the future. The first, *Great Present, Okay Future* (or GPOF for short), is a simplified version of what we have if we spend additional resources making present, existing lives better. Some number m of present people $(p_1, p_2, p_3, \ ...)$ have very great lives, and some much greater number n of future, contingent people $(f_1, f_2, f_3, \ ...)$ have lives that are merely okay but still worth living. GPOF can be illustrated as below, with each person's lifetime well-being corresponding to the height of their vertical line—the higher well-being levels here represent great lives, while the lower ones represent merely okay lives. (The presence of just two well-being levels is unrealistic, but note that the various arguments below go through with or without this assumption.)

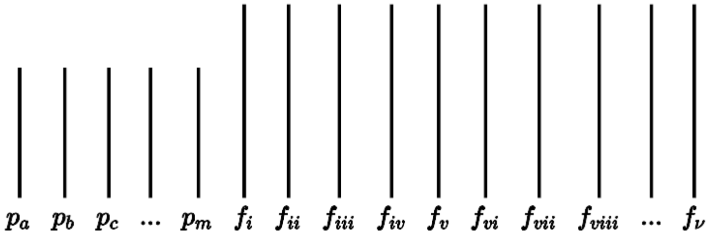


Great Present, Okay Future

The second such outcome, *Okay Present, Great Future* (OPGF), is an outcome we might obtain by putting additional resources towards making future, contingent lives better than those in GPOF, and succeeding in doing so. In this outcome, the same m present people $(p_1, p_2, p_3, \ ...)$ exist and have merely okay lives. And some much greater number of future, contingent people $(f_i, f_{ii}, f_{iii}, \ ...)$ have great lives. But now, for reasons given above, those future people are *different* from those who would exist in GPOF. And there might be a different total number of them, perhaps ν.[10]

[10] Astute readers may recognise the comparison between GPOF and OPGF as an analogue of Derek Parfit's case of *Depletion*. Readers who would like a more concrete case to attach to the comparison of GPOF and OPGF may keep the following in mind.

> **Depletion**
> As a community, we must choose whether to deplete or conserve certain kinds of resources. If we choose Depletion, the quality of life over the next two centuries would be slightly higher than it would have been



Okay Present, Great Future

Many of those who wish to deny In-Principle Longtermism will also want to deny the verdict that some such OPGF is better than some such GPOF. (Call this claim OPGF$\succ$GPOF for short.) As we'll see in later sections, it is extremely hard to accept this claim yet deny In-Principle Longtermism. Because of this, even though OPGF$\succ$GPOF isn't strictly necessary nor sufficient for In-Principle Longtermism, it certainly makes the case for it much stronger.

There is also a compelling independent reason for denying that OPGF$\succ$GPOF (to which both Setiya, 2022 and Plant, 2023 explicitly appeal): namely, that GPOF is better for some and *worse for no one*. Each present or near-future person is better off under GPOF than under OPGF. And no future person who has a merely okay life under GPOF would exist otherwise, so they have not been made worse off than they would otherwise be. And we might think that one outcome cannot be worse than another unless it is worse *for someone*. This reason too might be offered against In-Principle Longtermism—we might think that no prospect benefiting the long-term future, P_{Future}, could be better than a prospect benefiting the present because it makes no one better off. Nonetheless, to deny longtermism on such grounds brings on serious challenges, as we will see in what follows.

general impossibility results for those who wish to deny OPGF$\succ$GPOF.

3 Formal framework

To streamline the discussion, I will adopt some basic formalism and formal axiological assumptions.

Let $\mathcal{O}$ denote the set of all (metaphysically) possible outcomes. I assume that this set will be infinite, as there are infinitely many possible persons (and, *a fortiori*, infinitely many combinations of them), but it may be that each outcome contains only finitely many such persons. I assume also that, for any finite subset of possible people and any mapping of those persons to well-being levels, $\mathcal{O}$ contains at least one outcome in which precisely that set of people exist and precisely those well-being

> if we had chosen Conservation. But it would later, for many centuries, be much lower than it would have been if we had chosen Conservation. This would be because, at the start of this period, people would have to find alternatives for the resources that we had depleted. (Parfit, 1984, pp. 361-2)

levels.[11]

In any given decision problem, let $\mathcal{S}$ denote a set of states of the world, exactly one of which can obtain. There is some probability measure on $\mathcal{S}$, independent of which option the agent chooses. An *option* can be represented as a function from $\mathcal{S}$ to $\mathcal{O}$. It will also often be useful to speak of an option's corresponding *prospect* (e.g., P_i): its corresponding probability measure on $\mathcal{O}$.

Since what is at stake here is how we compare outcomes, options, and prospects, we need a betterness relation. Let $\succcurlyeq$ be a binary relation denoting 'is at least as good as' (with $\succ$ denoting 'is strictly better than' and $\sim$ denoting 'equally as good as'). I assume that this relation is defined on each of $\mathcal{O}$ and the set of (metaphysically possible) options, and that it is consistent between the two—that is, that outcomes $O_a \succcurlyeq O_b$ if and only if the option giving O_a in all states is at least as good as that which gives O_b in all states. I assume also that $\succcurlyeq$ is reflexive and non-symmetric, but not necessarily transitive nor complete on either set.

4 Impossibilities involving OPGF$\succ$GPOF

For those who find it implausible that it can be better to improve the long-term future than to improve the present, a variety of nasty impossibility results arise. In this section, I raise three such results which include among their jointly inconsistent conditions the denial that OPGF$\succ$GPOF (for at least some n, ν). For those who deny longtermism on the basis of their population-ethical views, these results will be troubling enough. For those who instead deny longtermism (at least partly) on the basis of their decision-theoretic views, discussion of risk and of In-Principle Longtermism itself will come later.

4.1 Average-Total-Equality Dominance

The first result is straightforward, and I won't dwell on it for long. It is simply that denying OPGF$\succ$GPOF is inconsistent with *Average-Total-Equality Dominance.*[12]

> *Average-Total-Equality Dominance*: If one possible outcome O_a contains a greater total sum of well-being, a greater average individual well-being, and less variance[13] in individual well-being than another outcome O_b, then $O_a \succ O_b$.

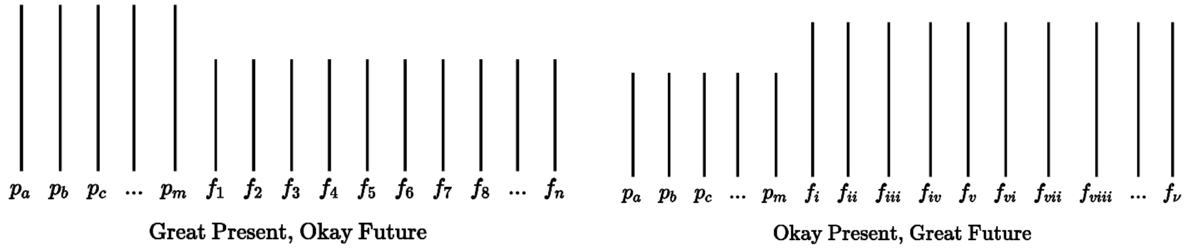
[11] These two assumptions are equivalent to what Thomas (2022b, p. 284) calls *Infinity* and *Rectangular Field.* They are necessary for the results in §4.3 and §5.3.

[12] Cf. the Non-Anti-Egalitarianism Principle used by, e.g., Arrhenius (2000, p. 253).

[13] Variance here is meant in the standard statistical sense: the population's average squared deviation from its average wellbeing. But we could replace it with other measures of equality in well-being—e.g., the Gini coefficient—and the below result would still hold.

Average-Total-Equality Dominance is plausible. In effect, it is the claim that moral comparisons of outcomes depend on nothing more than their total well-being, average well-being, and their levels of equality. For it not to hold, comparisons must depend on some further factor.

To see the inconsistency of this principle with the denial of OPGF$\succ$GPOF, consider again both distributions. We can easily observe that OPGF bests GPOF in all three respects: total well-being, average well-being, and equality.



Great Present, Okay Future **Okay Present, Great Future**

Consider the outcomes' total well-being. As long as the number ν of people in the Great Future is at least as great as the number m of present people (and not too much less than the number n of people in the Okay Future), then the total sum of well-being in OPGF is greater than that in GPOF.

Consider the outcomes' average well-being. For *any* $n, \nu > m$, OPGF will have greater average well-being per person.

And consider the outcomes' variance in individual well-being. As long as ν is large enough, OPGF will have a more equal distribution of well-being as well. As ν tends to infinity, the proportion of OPGF's population who have lives that deviate from 'great' approaches 0. And for any n, the proportion of GPOF's population who deviate from 'okay' will be some non-zero fraction. So, for any n, let ν be large enough and OPGF will have less variation in well-being than GPOF.

This suffices to show that Average-Total-Equality Dominance is inconsistent with denying *in full generality* that OPGF$\succ$GPOF. Even if a particular Great Future is not great enough that OPGF$\succ$GPOF, this dominance principle establishes that at least *some* such Great Future is great enough.

But how great a cost is it to deny Average-Total-Equality Dominance? The principle is upheld by many theories—e.g., totalism, averageism, various forms of egalitarianism—but not all. For instance, total prioritarianism violates it (see Holtug, 2007, §3), as do some person-affecting views (e.g., that of Bader, 2022). So, perhaps violating this principle, by itself, is not so great a cost to deny longtermism. But, as we will see below, there are further costs too.

4.2 Anonymity and Pareto

The next result is perhaps no less straightforward than the previous one, but I include it here for completeness. It is that the following three conditions, together with the denial that OPGF$\succ$GPOF, are jointly inconsistent.[14]

> *Anonymity*: If there is a bijection σ from the set of persons who exist in outcome O_a to the set of those who exist in outcome O_b such that each person is mapped to someone with precisely the same well-being in their respective outcome, then $O_a \sim O_b$.

Anonymity is the claim that the exact identities of the people within it make no difference to the goodness of an outcome—that any one outcome with such and such people at such and such well-being levels is no better nor worse than another with the same number of people at the same levels. And, on the face of it, this is an attractive principle. [15] It has been defended at length by various philosophers (e.g. Sidgwick, 1907; Parfit, 1984, §121 & Appendix F). And, even without such defences, it has some intuitive force behind it—if any account of overall moral goodness is to be plausible, it seems hard for it to say that it is better that Alice obtain high well-being than that Bob does, merely by virtue of their identities.

> *Same-Person Pareto*[16]: If precisely the same persons exist in outcomes O_a and O_b, and every such person has well-being at least as high in O_a as in O_b, then $O_a \succcurlyeq O_b$. If, as well, some person has strictly higher well-being in O_a than in O_b, then $O_a \succ O_b$.

This rather weak version of the Pareto principle tells us that, if two outcomes contain exactly the same people at exactly the same well-being levels, they are equally good; and, if were to then make one or more of those people better off and leave all else unchanged, then we would make the outcome better. At its most basic, this is the principle that what matters is how good each outcome is *for people*. And, again, this is hard to deny, unless we believe for independent reasons that something other than well-being matters for assessments of moral goodness. It is particularly hard to deny given that it applies *only* in comparisons where exactly the same people exist in both outcomes; it falls silent in comparisons where we may add additional people, or swap out some people for others.[17] As

[14] This result takes inspiration from the discussions of impartiality in Broome (2004, p. 136) and Bader (n.d., ch. 3). Cf. Parfit's (1984, §125) discussion of the *No-Difference View*.

[15] If we included in $\mathcal{O}$ outcomes in which *infinitely* many persons exist then Anonymity, as stated here, would be inconsistent with Same-Person Pareto (see Van Liedekerke, 1995; Wilkinson, 2021, p. 1929). But even then, we could restrict Anonymity to holding only for bijections which map only finitely many persons to different persons, and the two principles would be consistent. And, indeed, such a restricted form of Anonymity would suffice for my purposes here.

[16] This is equivalent to what Broome (2004, p. 120) calls the *Principle of Personal Good*.

[17] This distinguishes it from a strong version of the Pareto principle endorsed by narrow person-affecting views: that, if one outcome is at least as good as another for *every person who exists in both*, then it is at least as good overall; and, if the former outcome is strictly better than the latter for at least one such person, then it is better overall. This principle is jointly incompatible with Anonymity and the transitivity of $\succcurlyeq$ (Broome, 2004, p. 136).

such, it makes no particularly controversial claims about population ethics.

> *Transitivity of* $\succcurlyeq$: For any possible outcomes O_a, O_b, O_c, if $O_a \succcurlyeq O_b$ and $O_b \succcurlyeq O_c$, then $O_a \succcurlyeq O_c$.

This final condition tells us that the relation of ‘is (morally) better than’ works much like other commonplace comparative relations such as ‘hotter than’, ‘taller than’, and so on. It is the claim that we cannot replace an outcome with a better one, and do so again and again, and end up with an outcome worse than the first. Some philosophers claim that such comparative relations are necessarily and self-evidently transitive (e.g. Broome, 2004, pp. 50-1); some defend it on different grounds (see Huemer, 2008; Nebel, 2018). And others still claim that it is false, due to the existence of some intuitive moral verdicts that together violate transitivity (Temkin, 2011, at various points throughout). Nonetheless, I take it as at least prima facie plausible that moral betterness is transitive, and that needing to reject it would constitute at least some cost.

The impossibility of these conditions all holding at once can be observed as follows.

First, consider again GPOF and OPGF. For now, I’ll simplify both outcomes to contain only one person in the present/near-future, and two people in the long-term future ($m = 1, n, \nu = 2$), but the same argument can be run for *any* specific $n = \nu > m$.

$$\begin{array}{rccccc} & p & f_1 & f_2 & f_i & f_{ii} \\ \text{Great Present, Okay Future}: & 1 & 0 & 0 & - & - \\ \text{Okay Present, Great Future}: & 0 & - & - & 1 & 1 \end{array}$$

Consider also:

$$\begin{array}{rccccc} & p & f_1 & f_2 & f_i & f_{ii} \\ \sigma(\text{OPGF}): & 1 & 1 & 0 & - & - \end{array}$$

Note that the well-being levels of those in OPGF can be rearranged to obtain σ(OPGF). So, by Anonymity, OPGF$\sim$ σ(OPGF). And precisely the same persons exist in σ(OPGF) and GPOF. So, Same-Person Pareto implies that σ(OPGF)$\succ$GPOF. If we combine both observations, then the transitivity of $\succcurlyeq$ implies that OPGF$\succ$GPOF.

Thus, we have our impossibility. It turns out that, if you want to deny that OPGF$\succ$GPOF, you must deny at least one of: Anonymity; Same-Person Pareto; and the transitivity of $\succcurlyeq$.

But perhaps the denier of longtermism won’t find this result so troubling. After all, it depends on Anonymity. It wouldn’t be entirely surprising for such a denier to also deny this principle: to deny that present and future lives are effectively interchangeable. Perhaps denying this, too, is no great cost in order to deny that OPGF$\succ$GPOF and, with it, In-Principle Longtermism.

4.3 Same-Number Comparability

For those who do not find the previous result so troubling, suppose we replace Anonymity with an even less controversial condition. This result depends only on Same-Person Pareto, the transitivity of $\succcurlyeq$, and the new addition of *Same-Number Comparability.* This new addition is straightforward: it merely says that outcomes containing the same number of people are comparable.[18]

> *Same-Number Comparability*: For any integer $n > 0$ and any pairs of outcomes O_a and O_b that each contain n persons, O_a and O_b are comparable. That is, $O_a \succcurlyeq O_b$ or $O_a \preccurlyeq O_b$ (or both).

Here's how these conditions give us an impossibility.

Again, take GPOF and OPGF. And suppose again that OPGF is *not* better than GPOF (and not just in this case, but in full generality). But now introduce a third, hypothetical outcome, called O_* (containing a hypothetical person, f_*).

	p	f_1	f_2	f_i	f_{ii}		f_*
Great Present, Okay Future :	1	0	0	–	–		–
Okay Present, Great Future :	0	–	–	1	1		–
O_* :	–	1	–	0	–	–	0

And consider how O_* compares to OPGF, with the persons listed in a somewhat different order.

	f_i	p	f_{ii}	f_1	f_*	f_b
Okay Present, Great Future :	1	0	1	–	–	–
O_* :	0	–	–	1	0	–

You might notice that this pair of outcomes looks a lot like that of GPOF and OPGF. We have one person, f_i, who exists across both outcomes but has a better life in the first and so fills the role of the 'present' person. And we have two disjoint populations of possible 'future' people (p and f_{ii} in the first outcome; f_1 and f_* in the second). But, relative to the future people in the earlier comparison, some of those 'future' people are now better off in the first option and some are worse off in the second option.

We might like to compare these outcomes as we did with the pair above. But it is not immediately clear that we can do so. We are no longer choosing between benefiting a present person who exists

[18] Note that this condition is far weaker than the common assumption that the betterness relation is *complete* (e.g. Broome, 2004, p. 22). Same-Number Comparability does not require that *every* pair of possible outcomes is comparable, nor does it bring on the various strong implications that completeness does (e.g., in Harsanyi, 1955).

no matter what (f_i) or to bring one or another group of future people into existence. Given that OPGF is the very same outcome as above, f_i does not really live in the present and p does not really live in the future. The decisions are not entirely analogous.

But recall Same-Person Pareto. It says that, if exactly the same people exist at the same well-being levels, then the outcome is equally good; no matter whether those people exist in the present, the future, or the past.[19] It implies that OPGF would be equally good if f_i lived in the present rather than the future, and likewise if p existed in the future. OPGF is equally as good as such an outcome that is rearranged in time, and so, we can simply treat it as though it were such an outcome.[20] Likewise, we can treat O_* as an outcome in which f_i lives in the present and the others live in the future.

Given that, we must then deny that $O_* \succ$OPGF. After all, if we rearrange the times at which the people in O_* and OPGF live, then these two outcomes very nearly satisfy the original definitions of GPOF and OPGF. Indeed, by Same-Person Pareto, O_* is an outcome strictly *worse* than one that satisfies the definition of OPGF, and the rearranged version of OPGF is strictly *better* than the corresponding one satisfying the definition of GPOF. And recall that we have supposed, in full generality, that OPGF$\succ$GPOF is false. So, by Same-Person Pareto and the transitivity of $\succcurlyeq$, it must also be false that $O_* \succ$OPGF.

We can draw a similar implication when we compare O_* to GPOF, again listing the persons in a different order.

	f_1	f_i	f_*	p	f_2	f_{ii}
O_* :	1	0	0	–	–	–
Great Present, Okay Future :	0	–	–	1	0	–

Again, this pair of outcomes looks a lot like the earlier pair of GPOF and OPGF. And again, thanks to Same-Person Pareto, they are each equivalent to outcomes in which f_1 lives in the present and the others would all live in the future. And again, such temporally rearranged versions of these two outcomes would satisfy the definitions of GPOF and OPGF, except that the second outcome here is strictly worse than GPOF. So again, if we deny in full generality that OPGF$\succ$GPOF, then Same-Person Pareto and the transitivity of $\succcurlyeq$ imply that GPOF$\succ$ O_* must be false too.

In effect, if we accept Same-Person Pareto and we deny that OPGF$\succ$GPOF in full generality, then: OPGF can't be better than GPOF; O_* can't be better than OPGF; and GPOF can't be better than O_*. But, also, each pair must at least be comparable, by Same-Number Comparability. It must be that OPGF is *worse* than GPOF, or that they are equally good (and likewise for the other two

[19] This observation originates from Cowen (1992).

[20] This is due also to the transitivity of $\succcurlyeq$.

pairs). So, we can infer that GPOF$\succcurlyeq$OPGF$\succcurlyeq$ O_* $\succcurlyeq$GPOF.

Given that sequence, and given the transitivity of $\succcurlyeq$, none of those comparisons can be of strict betterness. After all, GPOF cannot be better than itself![21] So, they must all be equally good. But, if they are equally good, then we can still set up an option that is better than all of them. Take the pair (O_*,GPOF) from above and improve the life of f_2 in GPOF from okay to great. Call this new outcome O_{**}. Since it contains the same people as GPOF, but one is now strictly better off, O_{**} is better than GPOF by Same-Person Pareto. And given that GPOF must be equally as good as O_*, this outcome O_{**} must be better than $O*$ too.

	f_1	f_i	f_*	p	f_2	f_{ii}
O_* :	1	0	0	–	–	–
O_{**} :	0	–	–	1	1	–

By Same-Person Pareto, this pair is equivalent to one in which we let f_1 be the one present person in both outcomes and the others all be future persons. And, with that, we have a contradiction. With the people in each rearranged temporally, O_* and O_{**} satisfy the definitions of GPOF and OPGF, respectively. So, if O_{**} is strictly better, then we have violated the earlier assumption that OPGF is no better than GPOF.

Thus, if you want to deny that OPGF$\succ$GPOF, in full generality, then you must deny at least one of: Same-Number Comparability, Same-Person Pareto, and the transitivity of $\succcurlyeq$. And, at least to me, each of these conditions seems far harder to deny than Average-Total-Equality Dominance or Anonymity from above. Now, it seems, one who denies that it can be better to improve the long-term future than to improve the present must pay quite a serious cost.

5 Impossibilities involving In-Principle Longtermism

In the results described so far, I have been considering only outcomes, not risky prospects. I have considered whether an outcome that greatly benefits the long-term future must be better than one that modestly benefits the near term; not whether even a *slight probability* of greatly benefiting the long-term future must be better than a sure benefit to the near term. In this section, I turn my attention to the latter.

[21] This is due to the *reflexivity* of the $\succcurlyeq$, an entirely uncontroversial assumption. To my knowledge, no philosopher or welfare economist has entertained the notion that moral betterness may be irreflexive.

5.1 Avoiding Timidity

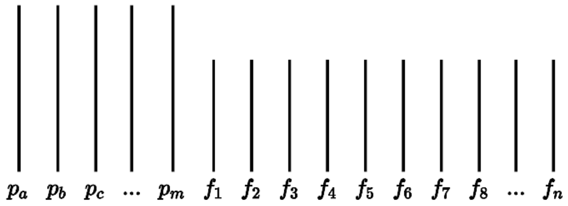
The first impossibility result in this new setting is that we cannot deny In-Principle Longtermism while also 1) accepting the transitivity of $\succcurlyeq$, 2) accepting that OPGF$\succ$GPOF (for some population sizes m, n, and ν), and 3) rejecting *Timidity*. Given the previous results that told us that OPGF$\succ$GPOF follows from various plausible combinations of principles, we could also redescribe this result such that the claim that OPGF$\succ$GPOF is replaced by, say, Same-Person Pareto and Same-Number Comparability.

> *Timidity*: There is some probability p such that, for *any* pair of prospects that can be described by P_p and $P_{p-\varepsilon}$, P_p is no worse than $P_{p-\varepsilon}$.
>
>> P_p: With probability p, some outcome O with finite population and finite total well-being; otherwise, some outcome O'.
>> $P_{p-\varepsilon}$: With probability $p-\varepsilon$, some outcome O^+ with the same population as O but *vastly* greater well-being for every person; otherwise, the same outcome O'.

Timidity says that it is sometimes no worse to pass up a prospect of arbitrarily large gains in well-being in order to reduce risk by even a tiny amount. The difference ε in probabilities might be as small as we like, perhaps 0.000001. And the better outcome O^+ can be as great an improvement over O as we like. Yet still $P_{p-0.000001}$ cannot be better than P_p if Timidity holds. This, you might agree, seems absurd.

Here is the argument that we cannot avoid Timidity without violating transitivity, denying that OPGF$\succ$GPOF, or accepting In-Principle Longtermism. It takes the form of a spectrum argument.[22]

Start with any given outcome described by GPOF or, equivalently, the prospect P_{Present} that gives that outcome for sure.

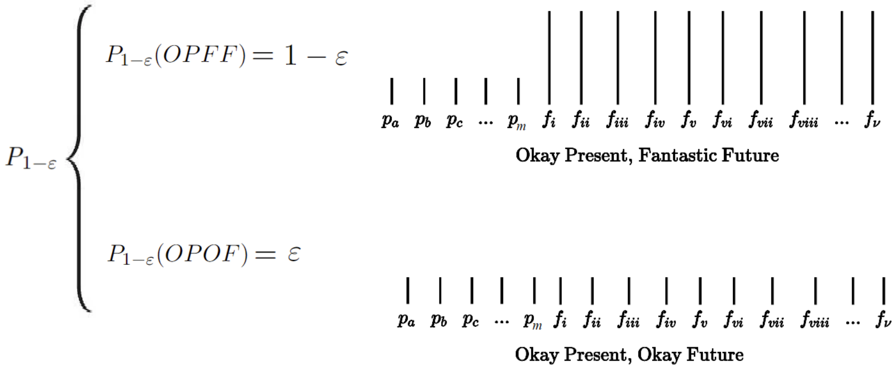


Great Present, Okay Future

We know from OPGF$\succ$GPOF (or, alternatively, from any of the conjunctions of conditions given in §3 above) that a prospect giving some outcome OPGF is even better. And we can then consider

[22] This argument, and the impossibility result it gives, is adapted from the titular result of Beckstead and Thomas (2021).

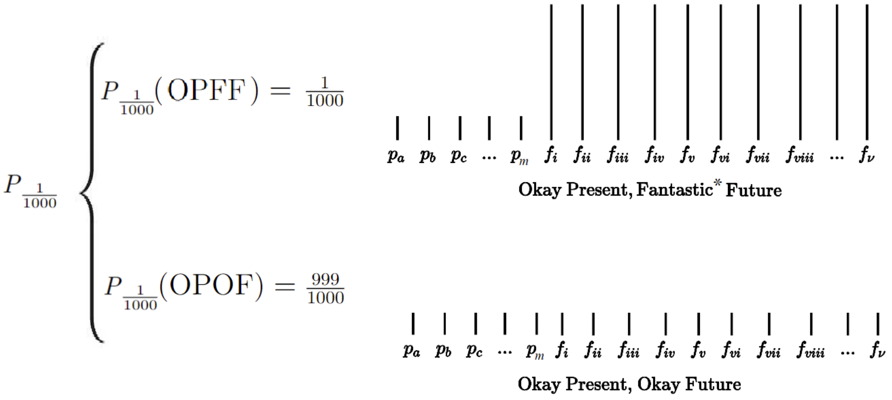
a prospect $P_{1-\varepsilon}$ that: 1) replaces OPGF with an outcome Okay Present, *Fantastic* Future which contains the same people but gives those future people *fantastic* lives; and 2) introduces a slight probability ε of an outcome Okay Present, Okay Future that gives those same people merely okay lives. We can represent such a prospect as follows.


$P_{1-\varepsilon}$
$P_{1-\varepsilon}(OPFF) = 1 - \varepsilon$
p_a p_b p_c ... p_m f_i f_{ii} f_{iii} f_{iv} f_v f_{vi} f_{vii} f_{viii} ... f_ν
Okay Present, Fantastic Future
$P_{1-\varepsilon}(OPOF) = \varepsilon$
p_a p_b p_c ... p_m f_i f_{ii} f_{iii} f_{iv} f_v f_{vi} f_{vii} f_{viii} ... f_ν
Okay Present, Okay Future

If we deny Timidity, then there is always some such $P_{1-\varepsilon}$ that is better than a prospect that gives OPGF for sure, which is in turn better than GPOF.

Likewise, we can reduce the probability of a great or fantastic future further, by ε each time. We can consider $P_{1-2\varepsilon}$, $P_{1-3\varepsilon}$, and so on, each an analogue of $P_{1-\varepsilon}$ but with the corresponding probability of some version of OPFF (with higher and higher well-being per person) and the complementary probability of OPOF. If we deny Timidity, then each is better than that which came before it.

Eventually, we have $P_{\frac{1}{1000}}$ (or $P_{1-n\varepsilon}$, if $\varepsilon = \frac{999}{1000n}$).


$P_{\frac{1}{1000}}$
$P_{\frac{1}{1000}}(\text{OPFF}) = \frac{1}{1000}$
p_a p_b p_c ... p_m f_i f_{ii} f_{iii} f_{iv} f_v f_{vi} f_{vii} f_{viii} ... f_ν
Okay Present, Fantastic* Future
$P_{\frac{1}{1000}}(\text{OPOF}) = \frac{999}{1000}$
p_a p_b p_c ... p_m f_i f_{ii} f_{iii} f_{iv} f_v f_{vi} f_{vii} f_{viii} ... f_ν
Okay Present, Okay Future

By the same reasoning, (at least some such) $P_{\frac{1}{1000}}$ must be better than the prospect that came before it, which must be better than the one that came before it, and so on all the way back to GPOF. Since $\succcurlyeq$ is transitive, then, $P_{\frac{1}{1000}}$ must be better than GPOF.

But this, you might notice, gives us In-Principle Longtermism. We have a possible pair of prospects P_{Present} and P_{Future} (GPOF and $P_{\frac{1}{1000}}$, respectively) such that P_{Present} is guaranteed to result in a better present, P_{Future} has probability no more than $\frac{1}{1000}$ of a better long-term future, and the sets of people existing in the long-term future in the two prospects are guaranteed to be entirely disjoint. So, if we deny In-Principle Longtermism, we have a contradiction.

To deny In-Principle Longtermism, we must therefore deny the transitivity of $\succcurlyeq$, or accept Timidity, or deny that OPGF$\succ$GPOF. And we know from above that denying the latter incurs the various costs detailed earlier, such as denying either Same-Person Pareto or both Same-Number Comparability and Anonymity.

5.2 Ex Ante Pareto

The next impossibility result replaces the denial of Timidity with the acceptance of both a weak form of ex ante Pareto and a principle about which options are better for individuals. We cannot deny In-Principle Longtermism while also accepting the transitivity of $\succcurlyeq$, that OPGF$\succ$GPOF (for some population sizes m, n, and ν), *Same-Person Ex Ante Pareto*, and *Personal Expectationalism.*

> *Same-Person Ex Ante Pareto*: If every possible outcome of options X and Y contains exactly the same persons and, for every such person a, X is ex ante at least as good for a as Y, then $X \succcurlyeq Y$. If, as well, for some such a, X is ex ante strictly better for a than Y, then $X \succ Y$.

This principle is *prima facie* extremely plausible. For any such options X and Y, there would be unanimous agreement among those affected that X is at least as good as (or strictly better than) Y. It would be deeply counterintuitive to then deny that X is at least as good as (or strictly better than) Y. Indeed, if we accept that moral betterness is grounded in what is better *for* persons,[23] this principle seems impossible to deny.

Easier to deny is the following principle, Personal Expectationalism: the view that what's best for a person is to have the greatest expectation of well-being. Many philosophers accept this, so denying it may constitute some cost. But, as I show below, we can weaken it greatly and still obtain the desired impossibility result.

> *Personal Expectationalism*: For any options X and Y and person a who exists in every possible outcome of both, X is ex ante at least as good for a as Y is iff X results in at least as great an expectation of (a particular cardinal measure of) a's well-being as does Y.

To see how these principles are together incompatible with the conditions listed above, consider

[23] This claim is often advocated by proponents of person-affecting views, e.g., Bader (2022, p. 260).

again two prospects from earlier: that which gives GPOF for sure; and $P_{\frac{1}{1000}}$, which has probability $\frac{1}{1000}$ of an outcome OPFF with a fantastic future and probability $\frac{999}{1000}$ of some outcome OPOF with a merely okay future. To deny In-Principle Longtermism, we must deny that $P_{\frac{1}{1000}}$ is better than GPOF (for *all* such GPOF, OPFF, and OPOF).

But we are assuming that OPGF$\succ$GPOF (for some such OPGF and GPOF), as well as the transitivity of $\succcurlyeq$. So, to deny In-Principle Longtermism, we must then also deny that $P_{\frac{1}{1000}}$ is better than OPGF; indeed, we must deny that it is better than OPGF even when the set of people in OPGF is precisely the same as in $P_{\frac{1}{1000}}$.

Consider both OPGF and $P_{\frac{1}{1000}}$ from the perspective of each person within them, as below.

OPGF

	1
p_1	okay
p_2	okay
...	okay
f_i	great
f_{ii}	great
...	great
f_μ	great

$P_{\frac{1}{1000}}$

	$\frac{1}{1000}$	$\frac{999}{1000}$
p_1	okay	okay
p_2	okay	okay
...	okay	okay
f_i	fantastic	okay
f_{ii}	fantastic	okay
...	fantastic	okay
f_μ	fantastic	okay

Whichever we choose of these two prospects, the same people will exist. And whichever we choose, each present person (p_1, p_2, ...) is guaranteed to have an okay life; for them, there is no difference. And each future person (f_i, f_{ii}, ...) faces the same two prospects individually: under OPGF, a great life, or, under $P_{\frac{1}{1000}}$, a fantastic life with probability $\frac{1}{1000}$ and an okay life otherwise. And we can assume that the difference between how good a fantastic life is for someone and how good a great life is for them is at least 999 times greater than the difference between how good a great life is and how good an okay life is (on whichever cardinal measure of well-being we like)—here, a fantastic life is especially fantastic.

Personal Expectationalism then applies here: the expected well-being of each person will then be greater in $P_{\frac{1}{1000}}$ than in OPGF, and so the former will be ex ante better for each person than the latter. And then Same-Person Ex Ante Pareto applies as well. These two options contain exactly the same possible people; indeed, exactly the same actual people. And for each such person $P_{\frac{1}{1000}}$ is ex ante strictly better than OPGF. So Same-Person Ex Ante Pareto tells us that $P_{\frac{1}{1000}} \succ$OPGF. And, by assumption, OPGF$\succ$GPOF. So, $P_{\frac{1}{1000}} \succ$GPOF. Thus, In-Principle Longtermism will hold.

So, to deny In-Principle Longtermism, we need to deny one of the following: Same-Person Ex Ante Pareto; Personal Expectationalism; that $\succcurlyeq$ is transitive; or that OPGF$\succ$GPOF (and, with it, each of the conjunctions of principles given in Section 3 above).

We can get a similar result even if we weaken Personal Expectationalism—even if we allow for risk-averse evaluations of ex ante betterness for individuals. Consider the principle of Timidity (from above), now adapted for individual persons.

> *Personal Timidity*: For some person a, there is some probability q such that, for *any* pair of prospects that can be described by P_p and $P_{p-\varepsilon}$, P_p is no worse than $P_{p-\varepsilon}$ for a.
>
> > P_p: With probability p, some outcome O in which a has finite well-being; otherwise, some outcome O'.
> >
> > $P_{p-\varepsilon}$: With probability $p-\varepsilon$, some outcome O^+ in which a has *vastly* greater well-being than in O; otherwise, the same outcome O'.

Much like in the interpersonal setting, Personal Timidity says that it is sometimes no worse to pass up arbitrarily large gains in well-being in order to reduce risk by even a tiny amount (even a difference in probability of 0.000000001). In this setting too, you may well agree that Timidity is absurd.

We can replace Personal Expectationalism with the denial of Personal Timidity and, by reasoning analogous to §4.1 above, there will be some $P_{\frac{1}{1000}}$ that is ex ante better for every single person than OPGF. The argument then works much the same: Same-Person Ex Ante Pareto then says that $P_{\frac{1}{1000}} \succ$OPGF; and, if OPGF$\succ$GPOF, then transitivity tells us that $P_{\frac{1}{1000}} \succ$GPOF; thus, we again have In-Principle Longtermism.

So, we cannot reject In-Principle Longtermism without rejecting one of: OPGF$\succ$GPOF; the transitivity of $\succcurlyeq$; Same-Person *Ex Ante* Pareto; and the denial of Personal Timidity. Again, the costs are mounting.

5.3 Separability

The next result no longer requires the condition that OPGF$\succ$GPOF. Instead, the conditions it finds to be jointly incompatible are: the denial of In-Principle Longtermism; the acceptance of Anonymity; and the acceptance of both *Prospect Separability*, the extremely weak principle of *Solipsist's Pareto*, and *Stochastic Dominance.*

> *Prospect Separability*: For any options X and Y, and any option Z for which the values of outcomes are probabilistically independent of both X and Y,[24] $X \succcurlyeq Y$ if and only if

[24] This condition—that the outcome of Z is probabilistically independent of those of both X and Y—is required because, without it, Prospect Separability would be incompatible with several other seemingly plausible principles. In particular, it is incompatible with the conjunction of Stochastic Dominance (defined below) and the transitivity of $\succcurlyeq$ (Russell, 2023; Wilkinson, n.d., §2). But the weaker version of Prospect Separability given here does not generate such problems (Wilkinson, n.d.).

$X \oplus Z \succcurlyeq Y \oplus Z.$

Here, $\oplus$ is a concatenation operation. Suppose A and B are outcomes. Then $A \oplus B$ is an outcome that features all of the (morally significant) events within A and all such events within B.[25] If we instead concatenate *options* like X and Z, then we concatenate their outcomes within each state. Equivalently, $X \oplus Y$ is the option that maps each state of the world s to the outcome $X(s) \oplus Y(s)$. In order to make Prospect Separability more plausible, we might also interpret $A \oplus B$ (or $X \oplus Y$) as outcomes (options) where the events taken from A (or X) are somehow isolated from the events taken from B (or Y). They might be causally isolated—the events taken from B might be placed outside the causal future of those from A, perhaps in the distant past of the events in A. Or the events from both outcomes/options might be placed outside the causal future of the other's.

Why accept Prospect Separability? Well, if we deny it, we fall prey to a version of the Egyptology Objection, which goes like this. Suppose you face a decision here and now in the twenty-first century, and the available options differ *only* in the events that occur in the future, and nearby in space. Your options don't differ in, say, what will have happened thousands of years ago in ancient Egypt (or, likewise, in distant galaxies). If so, then it seems that a moral comparison of your options then shouldn't depend on what happened there. That would be absurd, as ethicists have long noted (McMahan, 1981: p. 115; Parfit, 1984: p. 420). It seems obvious that matters of Egyptology are irrelevant to present-day decision-making, as is the study of other events that our actions do not affect. And yet, if we deny Prospect Separability, they will sometimes be: X and Y might represent your available options characterised in terms of their effects on the present and future, and Z some very sparse outcome containing only some events in that unaffected location (e.g., in ancient Egypt); if X is at least good as Y, then admitting that $X \oplus Z$ is *not* at least as good as $Y \oplus Z$ is allowing your comparison to be sensitive to unaffected, unrelated events in, say, ancient Egypt.

Consider also the eminently plausible principle of *Solipsist's Pareto*.

> *Solipsist's Pareto*: If two outcomes O_a and O_b both contain *only* one person p, $O_a \succcurlyeq O_b$ if and only if p's well-being is at least as great in O_a as in O_b.

It turns out that Prospect Separability and Solipsist's Pareto together have far-reaching implications. These two principles, in conjunction with Anonymity (from §3.2 above), imply an *additive* theory of moral betterness: they imply that one outcome is at least as good as another if and only

[25] Note that many location-specific events will inevitably be incompatible: e.g., the event "Nelson died at Trafalgar in 1805" and the event "Nelson survived Trafalgar in 1805". So, in concatenating sets of such events, we must do away with some of their morally insignificant features, such as the exact time and place they occurred. For instance, if outcome A contains Nelson dying at Trafalgar in 1805 and another outcome B contains him surviving, then we might construct $A \oplus B$ such that one man qualitatively identical to Nelson dies in a sea battle on one planet at a time called 1805 by his contemporaries and another man qualitatively identical to Nelson survives a similar sea battle on another planet at a time also called 1805 by his contemporaries.

if it contains at least as great a total sum (on some commutative, associative, and invertible notion of addition) of (some measure of) the well-being of each person that exists therein. (For a proof and relevant discussion, see Thomas, 2022b.)[26] This leaves open the question of whether adding additional lives is ever an improvement (which will be important below), how well-being is measured for the purpose of summation, and whether its sum is real-valued or is so complicated that it must be represented in some other way. But it does tell us this: an outcome like OPGF is better than an outcome like GPOF, where both contain populations of the same size; an outcome with a few of people with great lives and many with merely okay lives is not as good as an outcome with the few people living okay lives and the many living great lives.

So, Prospect Separability, Solipsist's Pareto, and Anonymity give us the result that OPGF$\succ$GPOF (for all $m < n = \mu$). But, if we add one more condition, we get In-Principle Longtermism too. That one remaining condition is *Stochastic Dominance.* Informally, this principle says that if two options give exactly the same probabilities of the same outcomes (or equally good outcomes) then they are equally good; and if you then swap out any of those outcomes in one option for even better ones, then you make the option strictly better. This seems overwhelmingly plausible. More formally, it can be stated as follows.

> *Stochastic Dominance*: Let X and Y be any two options. If, for every possible outcome $O \in \mathcal{O}$, X has at least as high a probability as Y of resulting in an outcome at least as good as O, then $X \succcurlyeq Y$.
>
> If, as well, there is some possible outcome $O \in \mathcal{O}$ such that X has strictly greater probability than Y of resulting in an outcome at least as good as O, then X is strictly better than Y.

Combine Stochastic Dominance with Prospect Separability, Solipsist's Pareto, and Anonymity and we have not only that OPGF$\succ$GPOF but that $P_{\frac{1}{1000}} \succ$GPOF too. Why? A more comprehensive proof appears in the appendix, but the basic argument goes as follows.

Let $u(X)$ be some real-valued utility function, representing $\succcurlyeq$, on some subset of possible outcomes containing OPOF, GPOF, OPFF, and some maximal collection of other outcomes that are comparable with them and that are mutually comparable. For now, let $u(X)$ be any *ordinal* utility function: $u(A) \geq u(B)$ if and only if $A \succcurlyeq B$, but the difference between $u(A)$ and $u(B)$ need not tell us anything about how much better A is than B. Given any such $u(A)$, we can describe $P_{\frac{1}{1000}}$ and GPOF graphically as cumulative probability distributions. (For any given value, a cumulative probability is simply a function that outputs that option's probability of resulting in a value less than or equal to that.)

[26] Similar results are proven by Blackorby et al. (2005, thm 6.10) and Pivato (2014, thm 1), among others.

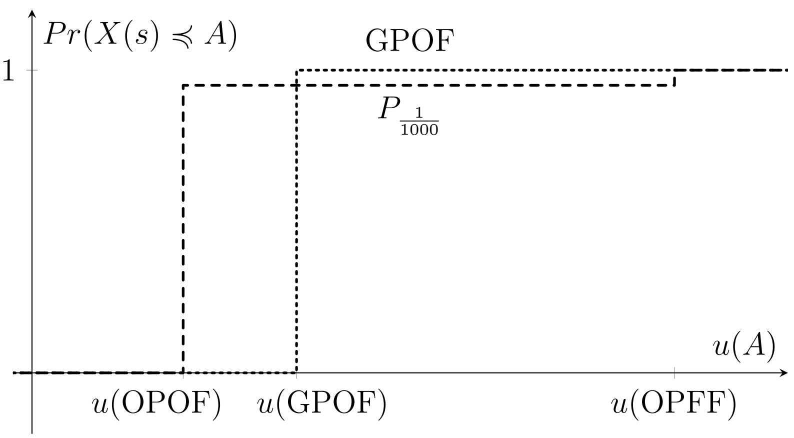


Figure 1: Cumulative probability distributions for $P_{\frac{1}{1000}}$ and GPOF

We can concatenate both options with a third prospect, Z. This prospect will have some non-zero probability density assigned to every utility. As utility tends to positive or negative infinity, Z's probability density will tend to zero but, crucially, it will do so only very slowly (more on this in the appendix). For at least some such Z (and some OPOF, GPOF, and OPFF), it can be shown that the concatenated options $P_{\frac{1}{1000}} \oplus Z$ and GPOF$\oplus Z$ will look as follows (see appendix).

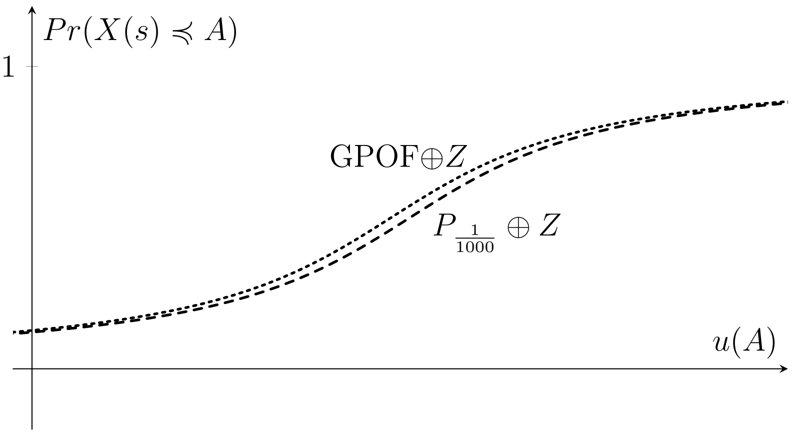


Figure 2: Cumulative probability graphs for $P_{\frac{1}{1000}} \oplus Z$ and GPOF$\oplus Z$

As can be seen here, the cumulative distribution for GPOF$\oplus Z$ is always at least as high as, and sometimes higher than, that for $P_{\frac{1}{1000}} \oplus Z$. In other words, for any possible value, the former is *no more* likely to result in more value than that than the other is. This means that, for every possible outcome, $P_{\frac{1}{1000}} \oplus Z$ is at least as likely to result in an outcome that good as GPOF$\oplus Z$. And, for some outcomes, $P_{\frac{1}{1000}} \oplus Z$ is strictly *more* likely to result in an outcome that good. If this relationship sounds familiar, that's because it's the condition for strict betterness that Stochastic Dominance gives us. According to Stochastic Dominance, $P_{\frac{1}{1000}} \oplus Z$ must be strictly better than GPOF$\oplus Z$.

But recall that, by Prospect Separability, $P_{\frac{1}{1000}} \oplus Z$ can only be better than GPOF$\oplus Z$ if $P_{\frac{1}{1000}}$

is better than GPOF too. So we cannot accept this assortment of principles without also accepting In-Principle Longtermism!

Thus, we have our final impossibility result. We cannot accept Prospect Separability, Stochastic Dominance, Anonymity, and Solipsist's Pareto while also rejecting In-Principle Longtermism.

6 Preventing extinction

The above results, by and large, were established using outcomes and prospects for which the verdicts of longtermism are especially plausible—outcomes and prospects between which we can increase the well-being of future people without changing the number of people who exist. For the most part, the versions of GPOF, OPGF, and other outcomes described above, as well as $P_{\frac{1}{1000}}$, all contain populations of the same size.

But this feature is unrealistic. In practice, actions with widespread effects on the well-being of future people will also affect the size of the future population. After all, the identities of future people are extremely sensitive to small changes and, as a result, so too will be the events within their lives, including their own choices of how many children to have.

In addition, as a practical matter, many who endorse longtermism happen to advocate for future-affecting actions that not only improve the well-being of future people but also reduce the risk of human extinction (e.g. Ord, 2020; Greaves and MacAskill, 2021; MacAskill, 2022, ch. 8). As just one concrete example, each of those same authors suggest that one promising way to benefit the long-term future is to implement policies that strengthen healthcare systems in such a way to reduce the risk of future global pandemics. Doing so would offer several possible benefits: 1) reducing the frequency of pandemics in general, and so straightforwardly increasing the well-being of many future people; 2) reducing the probability of extremely disruptive pandemics that could lead to the collapse of institutions and entire nations, and so reducing risks of long-lasting reductions in well-being; and 3) reducing the risk of those most lethal pandemics that might lead to the wholesale extinction of humanity. But are these supposed benefits, in combination, really so great? The results presented above do not tell us, as they seem to only apply to situations where affecting extinction risk is *not* on the table—where the size of the future population is the same no matter what.

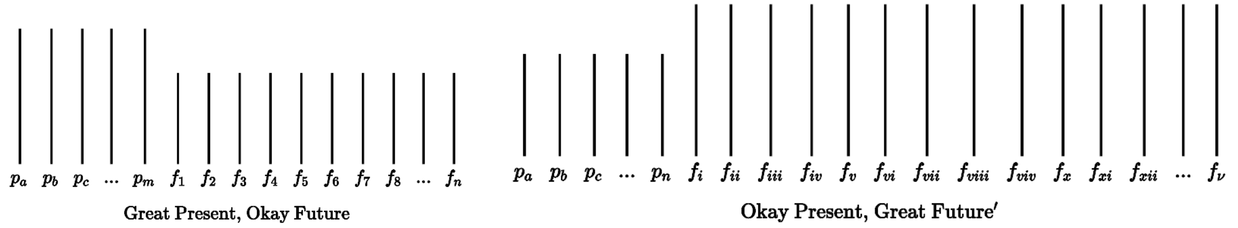
Could we still deny that it can be best to benefit the long-term future if, in practice, doing so always changes the population size—if doing so is accompanied by a reduction in the risk of human extinction? Could we deny In-Principle Longtermism even if the P_{Future} in the definition had a greater future population in either or both of its outcomes? The answer turns out to be no, at least not without incurring costs similar to those detailed above. It turns out that several of the above results also apply even if an improved future must inevitably be a more populous future. And the remaining results apply too, if we modify them slightly.

In what follows, for any given GPOF, I'll denote by OPGF$^+$ an outcome with greater future population than GPOF (i.e., with $\nu > n$). And I'll denote by $P^+_{\frac{1}{1000}}$ a corresponding prospect with probability $\frac{1}{1000}$ of some OPFF$^+$ with a much larger future population than GPOF, and probability $\frac{999}{1000}$ of some outcome OPOF$^+$ which may or may not have a larger future population than GPOF.

6.1 Average-Total-Equality Dominance

Recall that Average-Total-Equality Dominance says that one outcome must be better than another if it has all of: a greater total sum of well-being; a greater average of individual well-being; and less variance in individual well-being. In §3.1 above, OPGF did better than GPOF in all three respects, so the principle implied that OPGF$\succ$GPOF.

But the result there did not depend on OPGF and GPOF having the same future population size. Indeed, we might replace OPGF with an outcome OPGF$^+$ which differs only in that contains vastly more future people with great lives. OPGF$^+$ will have an even greater total sum of well-being, an even greater average individual well-being, and even less variance in individual well-being.


p_a p_b p_c ... p_m f_1 f_2 f_3 f_4 f_5 f_6 f_7 f_8 ... f_n
p_a p_b p_c ... p_n f_i f_{ii} f_{iii} f_{iv} f_v f_{vi} f_{vii} f_{viii} f_{viv} f_x f_{xi} f_{xii} ... f_ν

Great Present, Okay Future

Okay Present, Great Future′

So, by Average-Total-Equality Dominance, the side effect of increasing the size of the future population does not undermine the claim that OPGF$^+$ $\succ$GPOF. Indeed, increasing the future population size can itself make the outcome better. Thus, we still have the analogous impossibility result: that Average-Total-Equality Dominance is incompatible with the wholesale denial that OPGF$^+$ $\succ$GPOF.

6.2 Anonymity, Pareto, and Benign Excellent Addition

Recall the result from §3.2: that Anonymity, Same-Person Pareto, the transitivity of $\succcurlyeq$, and the denial of OPGF$\succ$GPOF are incompatible. To extend this result to replace that last condition with the denial of OPGF$^+$ $\succ$ GPOF, we merely need to add the condition of Benign Excellent Addition.

> *Benign Excellent Addition*: For some sufficiently high well-being w, if an outcome O_b contains all of the same persons as O_a and just one additional person p with well-being

> w, and if O_b would be strictly better than O_a if just p were removed, then O_b is at least as good as O_a.

This addition does not seem a controversial one. It seems that adding an additional person to a population, at least if that person is extremely well-off, must not make that population worse. It may not make the population better, but making it worse would be counterintuitive.[27] And if the addition of the extremely well-off person is accompanied by other well-being gains among the existing population, it certainly seems that the resulting outcome cannot be worse.

It is straightforward to show that this condition, when combined with those listed above, gives us an impossibility. By the same reasoning as in §3.2 above, Anonymity, Same-Person Pareto, and the transitivity of $\succcurlyeq$ together imply that OPGF$^-$ below is better than GPOF.

	p	f_1	f_2	f_i	f_{ii}	f_{iii}
GPOF :	1	0	0	–	–	–
OPGF$^-$:	0	–	–	1	0.5	–
OPGF$^+$:	0	–	–	1	1	1

Benign Excellent Addition, when combined with Same-Person Pareto, gives us the further step that the above OPGF$^+$ is better than OPGF$^-$. After all, everyone in OPGF$^-$ is at least as well off in OPGF$^+$ and one person is strictly better off, so Same-Person Pareto says that OPGF$^+$ with person f_{iii} omitted would be better. And then we can define a great life as one with well-being of at least w, so Benign Excellent Addition would then say that OPGF$^+$ $\succ$OGPF$^-$. Then, transitivity implies that OPGF$^+$ $\succ$GPOF. (Indeed, the same would hold for GPOF and OGPF$^+$ with any population sizes $n > \nu > m > 0$.)

Thus, we have the following revised result. If you want to deny that OPGF$^+$ $\succ$GPOF (for all $\nu > n > m > 0$), you must deny at least one of: Anonymity; Same-Person Pareto; the transitivity of $\succcurlyeq$; and Benign Excellent Addition.

6.3 Same-Number Comparability and Benign Excellent Addition

Much like the previous result, to extend the result from §3.3, we again merely need to add Benign Excellent Addition.[28]

[27] One reason to be suspicious of Benign Excellent Addition is that a similar principle (merely *Benign Addition* or *Mere Addition*), combined with a principle like Average-Total-Equality Dominance, implies the much-maligned Repugnant Conclusion. But it is worth noting that Benign *Excellent* Addition does not do so, nor does the conjunction of it and the other principles listed here.

[28] In fact, for this result, we can use a strictly weaker version of the principle: *Weak Benign Excellent Addition.*

> *Weak Benign Excellent Addition*: For some sufficiently high well-being w, if an outcome O_b contains all of the same persons as O_a and just one additional person p with well-being w, and if O_b would be strictly

By the same reasoning as in §3.3 above, Same-Number Comparability, Same-Person Pareto, and the transitivity of $\succcurlyeq$ together imply that OPGF$^-$ is strictly better than the GPOF given below. And Benign Excellent Addition implies that OPGF$^+$ is at least as good as OPGF$^-$. So, by transitivity, OPGF$^+$ $\succ$GPOF. (And, likewise, the same argument can be run for any $\nu > n > m > 0$.)

	p	f_1	f_2	f_i	f_{ii}	f_{iii}
GPOF :	1	0	0	–	–	–
OPGF$^-$:	0	–	–	1	0.5	–
OPGF$^+$:	0	–	–	1	1	1

Thus, we have an impossibility: Same-Number Comparability, Same-Person Pareto, the transitivity of $\succcurlyeq$, Benign Excellent Addition, and denial that OPGF$^+$ $\succ$GPOF (for all $\nu > m, n > 0$) are incompatible.

6.4 Avoiding Timidity

We can update the result from §4.1 as follows. We cannot reject In-Principle Longtermism even when improving the future requires greatly expanding the future population (and so, as a necessary condition, reject that $P^+_{\frac{1}{1000}}$ $\succ$GPOF) without also: accepting Timidity; denying the transitivity of $\succcurlyeq$; or denying that OPGF$^+$ $\succ$GPOF (for all OPGF$^+$ and GPOF).

The argument for this result is much like that from earlier. From OPGF$^+$ $\succ$GPOF, we know that some OPGF$^+$ containing ν future people is better than GPOF containing $n < \mu$ future people. Then, take some prospect $P^+_{1-\varepsilon}$ with probability $1 - \varepsilon$ of some much better OPFF$^+$ and probability ε of some worse OPOF$^+$ (each with the same number of future people as OPGF$^+$). Unless Timidity holds, some such $P^+_{1-\varepsilon}$ must be better than OPGF$^+$. Likewise, unless Timidity holds, some $P^+_{1-2\varepsilon}$ must be better still, and some $P^+_{1-3\varepsilon}$ better still than that, and so on until we reach some $P^+_{\frac{1}{1000}}$. By transitivity, that $P^+_{\frac{1}{1000}}$ must also be better than OPGF$^+$. And, again by transitivity, it must be better than GPOF too, despite $P^+_{\frac{1}{1000}}$ having a higher population.

So, we have the result that we cannot deny that $P^+_{\frac{1}{1000}}$ $\succ$GPOF unless we also accept Timidity, deny transitivity, or deny that OPGF$^+$ $\succ$GPOF (and, along with it, deny one of the conjunctions of principles given above).

better than O_a if just p were removed, then O_b is *no worse than* O_a.

This version, by itself, is compatible with such an O_a and O_b being *incomparable*, rather than O_b needing to be at least as good as O_a.

6.5 Ex Ante Pareto

Like the previous result, we can run much the same argument as before, but with the condition that OPGF$\succ$GPOF replaced with OPGF$^+$ $\succ$GPOF.

Since OPGF$^+$ $\succ$GPOF and since $\succcurlyeq$ is transitive, to deny that $P^+_{\frac{1}{1000}}$ $\succ$GPOF we must also deny that $P^+_{\frac{1}{1000}}$ $\succ$OPGF$^+$. But then we can compare a version of OPGF$^+$ to a version of $P^+_{\frac{1}{1000}}$ containing exactly the same future population—the same number of persons, and exactly the same persons as well. And as we saw in §4.2, for such a pair of options, Same-Person Ex Ante Pareto and the denial of Personal Timidity together imply that $P^+_{\frac{1}{1000}}$ $\succ$OPGF$^+$.

Thus, we have the result that we cannot deny that $P^+_{\frac{1}{1000}}$ $\succ$GPOF unless we also deny Same-Person Ex Ante Pareto, accept Personal Timidity, deny transitivity, or deny that OPGF$^+$ $\succ$GPOF (and, along with it, deny one of the conjunctions of principles given above).

6.6 Separability

The final result from earlier extends neatly as well, with the mere additions of Benign Excellent Addition and the transitivity of $\succcurlyeq$. The earlier result was that Prospect Separability, Stochastic Dominance, Anonymity, and Solipsist's Pareto are together incompatible with rejecting In-Principle Longtermism. But, together with Benign Excellent Addition and transitivity, they are also incompatible with rejecting In-Principle Longtermism when improving the future inevitably means growing the future population—incompatible with denying that $P^+_{\frac{1}{1000}}$ $\succ$GPOF.

From the result above, we know that these principles imply that some version of $P_{\frac{1}{1000}}$ is strictly better than GPOF, where the outcomes of $P_{\frac{1}{1000}}$—OPOF and OPGF—have present populations and future populations each of the same size, respectively, as those in GPOF. But we can swap out the outcome OPGF for OPGF$^+$, with a larger and even better-off future population, and Benign Excellent Addition will say that it is at least as good. The resulting prospect $P^+_{\frac{1}{1000}}$ will then be at least as good as $P_{\frac{1}{1000}}$, by Stochastic Dominance. And, since $P_{\frac{1}{1000}}$ was already better than GPOF, by the transitivity of $\succcurlyeq$, we can conclude that $P^+_{\frac{1}{1000}}$ is better than it too, as required.

Thus, we cannot avoid this verdict unless we deny at least one of: Prospect Separability; Solipsist's Pareto; Anonymity; Stochastic Dominance; Benign Excellent Addition; and the transitivity of $\succcurlyeq$.

7 Conclusion

Longtermism may seem a bold and counterintuitive moral claim. It may seem preposterous that it could be better to benefit future unborn people than to benefit our contemporaries, especially when we can only do so with low probability and only while also changing the identities of those supposed beneficiaries. This may seem like an idea that only a total utilitarian would be foolish enough to accept.

But, as we have seen, it is not so easy to deny longtermism. At least, it is not so easy to deny it in the manner that many critics do, without any need to examine unobvious descriptive facts—to deny that it would hold even if our descriptive circumstances were favourable to it, if the future were large enough and if we had options that presented at least a slight probability of improving that future. To deny longtermism in this way, we must pay some, arguably troubling, costs.

To start with, to deny that it can be better to benefit future people than present people when risk is absent, we must also deny several conjunctions of plausible principles. We must deny Average-Total-Equality Dominance (from §3.1). We must deny Anonymity or Same-Person Pareto or the transitivity of moral betterness (from §3.2). And we must deny Same-Number Comparability or, again, Same-Person Pareto or transitivity (from §3.3).

Further, to deny In-Principle Longtermism, we must deny either the aforementioned verdict or some further, highly plausible principles. We must accept Timidity or deny transitivity or deny that aforementioned risk-free verdict (from §4.1), and with it deny the conjunctions of principles listed above. In addition, we must deny Same-Person Ex Ante Pareto, or accept Personal Timidity, or again deny the aforementioned risk-free verdict (from §4.2). And we must deny Prospect Separability or Solipsist's Pareto or Anonymity or Stochastic Dominance.

For the most part, these results still hold if we make the cases in question more realistic—in particular, if we restrict the possible means of improving future lives to ones that also increase the number of those future lives. With this condition, for risk-free cases: the first result still holds; while, for the second and third, we need only add Benign Excellent Addition to the list of conditions for the results to hold. And then, for risky cases, all three of the analogous results hold.

Thus, we find ourselves between a rock and numerous hard places if we wish to deny longtermism without leaving the armchair. And we find ourselves in a similar position if we wish to similarly deny that it can be valuable to mitigate risks of human extinction. To do so, we must incur various counterintuitive costs. In my view, these costs are collectively too great to bear. Perhaps, instead, we should simply accept longtermism. Or at least, if we deny it, we must do so by appeal to (perhaps contentious) descriptive facts.[29]

[29] Such an approach is taken by, for instance, Thorstad (2023a,b).

8 Appendix A: Proof of result from §4.3

Following Thomas (2022b, pp. 289-290), we can show that Prospect Separability, Solipsist's Pareto, and Anonymity together imply a totalist, additive theory of moral betterness as follows.

Recall that $\mathcal{O}$ is the set of outcomes. For any outcome $O \in \mathcal{O}$, denote the equivalence class of O by $[O]$. Let $\mathcal{O}_e$ be the set of all such equivalence classes, and let $\geq$ be the partial order on $\mathcal{O}_e$ given by: for all $O_a, O_b \in \mathcal{O}$, $[O_a] \geq [O_b]$ if and only if $O_a \succcurlyeq O_b$.

We can then define addition on $\mathcal{O}_e$, with $[O_a] + [O_b]$ equal to $[O_a \oplus O_b]$ for all $O_a, O_b \in \mathcal{O}$. If we accept Anonymity and Prospect Separability then this operation will be commutative and associative, and tells us that $[O_a] + [O_c] \geq [O_b] + [O_c]$ if and only if $[O_a] \geq [O_b]$ (Thomas, 2022b, p. 290). We can also let O_0 be an outcome containing no (morally valuable) events, in which case $[O_a] + [O_0] = [O_a \oplus O_0] = [O_a]$ for all $O_a \in \mathcal{O}$.

We can also define a subtraction operation such that $[O_a] - [O_b] \geq [O_c] - [O_d]$ if and only if $[O_a] + [O_d] \geq [O_b] + [O_c]$. And, with this, we can define $\mathcal{U}$ as the set of all differences $[O_a] - [O_b]$ for $O_a, O_b \in \mathcal{O}$. Each equivalence class $[O]$ in $\mathcal{O}_e$ will correspond to the difference $[O] - [O_0]$ in $\mathcal{U}$. (We can simply write $[O]$ for both the equivalence class and the corresponding difference.) We can then define $\geq$ and $+$ in $\mathcal{U}$ as in $\mathcal{O}_e$, with the same properties.

This is enough to make $\mathcal{U}$ a partially ordered abelian group. And, for such a group, we can represent the $\geq$ relation (and thereby the $\succcurlyeq$ relation) with some utility function $u([O])$, such that $u([O_a] \pm [O_b]) = u([O_a]) \pm u([O_b])$ for all $O_a, O_b \in \mathcal{O}$ (for some commutative, associative addition operation). Indeed, this utility function will be essentially unique—all such utility functions will agree on which sums and differences of (equivalence classes of) outcomes are greater than others. But note that this utility function need not be real-valued—for instance, it need not satisfy the Archimedean Condition that, for any $[O_a], [O_b] \geq [O_0]$, there is always some natural number n such that $n \times u([O_a]) \geq u([O_b])$ (where scalar multiplication is defined via repeated addition).

It follows from $u([O_a] \pm [O_b]) = u([O_a]) \pm u([O_b])$ (for all $O_a, O_b \in \mathcal{O}$) that the utility of an outcome containing n persons is the total sum of the utility of n outcomes that each contain one of those persons at the same well-being level. And it follows from Solipsist's Pareto that, if we take any two outcomes O_c and O_d that each contain only one person, $u([O_c]) \geq u([O_d])$ if and only if the person in O_c has well-being at least as great as has the person in O_d. Combining these two points, we have the result that, for any outcomes $O_a, O_b \in \mathcal{O}$, it will hold that $O_a \geq O_b$ if and only if the total sum of (a particular measure of) the well-being of each person in O_a is at least as great as the corresponding total sum for O_b. In short, Prospect Separability, Solipsist's Pareto, and Anonymity together imply a totalist, additive theory of moral betterness.

With this result in hand, we can then show that there is some prospect Z such that $P_{\frac{1}{1000}} \oplus Z \succ$GPOF$\oplus Z$ (for some such $P_{\frac{1}{1000}}$ and GPOF), by Stochastic Dominance (cf. Tarsney, 2020).

First, let GPOF, OPGF, and OPOF be defined such that $n = \nu \geq 1001m > 0$. In other words, the number of future people in each of the three outcomes is at least 1,001 times the number of present people in each outcome. Then, let $P_{\frac{1}{1000}}$ be the prospect that gives OPGF with probability 0.001 and OPOF with probability 0.999.

Then, let G be an outcome containing one person with a great life (at whatever level of well-being we define as being great) and O be an outcome containing one person with a merely okay life. By the properties of the utility function given above,

$$u([\text{OPGF}]) = m \times u([O]) + n \times u([G]),$$

$$u([\text{GPOF}]) = m \times u([G]) + n \times u([O]), \text{ and}$$
$$u([\text{OPOF}]) = (m + n) \times u([O])$$

Next, let Z be a prospect, probabilistically independent of $P_{\frac{1}{1000}}$, that results in outcome $a \times [G] - a \times [O]$ with probability density $p(a) = \frac{1}{2s} e^{\frac{|a|}{s}}$ for every real a and for some fixed $s > 1,000,000m$. This Z is chosen such that its probability distribution is a Laplace distribution with scale factor s and centred on O_0. Given the properties of a Laplace distribution, this means that the cumulative probability distribution of Z satisfies

$$Pr\big([Z(s)] \leq a[G] - a[O]\big) = \begin{cases} \frac{1}{2} e^{\frac{a}{s}} \text{ for } a \leq 0 \\ 1 - \frac{1}{2} e^{\frac{a}{s}} \text{ for } a > 0 \end{cases} \tag{1}$$

Take any outcome $O_b \in \mathcal{O}$ that satisfies $[O_b] = b[G] + (m + n - b)[O]$ for some real b. Then the probability that GPOF$\oplus Z$ results in an outcome GPOF$\oplus Z(s)$ at least as good as O_b is

$$\begin{aligned} Pr\Big(\text{GPOF} \oplus Z(s) \succcurlyeq O_b\Big) &= Pr\Big(m[G] + n[O] + a[G] - a[O] \geq b[G] + (m + n - b)[O]\Big) \\ &= Pr\Big(a[G] - a[O] \geq (b - m)[G] - (b - m)[O]\Big) \\ &= Pr\big(a \geq b - m\big) \\ &= 1 - Pr\big(a \leq b - m\big) \end{aligned} \tag{2}$$

And the probability that $P_{\frac{1}{1000}} \oplus Z$ results in an outcome $P_{\frac{1}{1000}}(s) \oplus Z(s)$ at least as good as O_b is

$$Pr\Big(P_{\frac{1}{1000}}(s) \oplus Z(s) \succcurlyeq O_b\Big) = \frac{1}{1000} Pr\Big(n[G] + m[O] + a[G] - a[O] \geq b[G] + (m + n - b)[O]\Big)$$

$$+\frac{999}{1000}Pr\Big((m+n)[O]+a[G]-a[O]\geq b[G]+(m+n-b)[O]\Big)$$

$$=\frac{1}{1000}Pr\Big(a[G]-a[O]\geq (b-n)[G]-(b-n)[O]\Big)+\frac{999}{1000}Pr\Big((a[G]-a[O]\geq b[G]-b[O]\Big)$$

$$=\frac{1}{1000}Pr\big(a\geq b-n\big)+\frac{999}{1000}Pr\big(a\geq b\big)$$

$$=\frac{1}{1000}\Big(1-Pr\big(a\leq b-n\big)\Big)+\frac{999}{1000}\Big(1-Pr\big(a\geq b\big)\Big)$$

$$=1-\frac{1}{1000}Pr\big(a\leq b-n\big)-\frac{999}{1000}Pr\big(a\geq b\big) \qquad (3)$$

We can combine (1), (2), and (3) to establish that (1) is strictly greater than (2) for all real b. To do so, consider four (mutually exclusive and exhaustive possibilities): 1) that $b<0$; 2) that $0\leq b<m$; 3) that $m\leq b<n$; and 4)that $n\leq b$.

1) for $b\leq 0$, (1) implies that the difference between (3) and (2) is

$$1-\frac{1}{1000}\big(\frac{1}{2}e^{\frac{b-n}{s}}\big)-\frac{999}{1000}\big(\frac{1}{2}e^{\frac{b}{s}}\big)-\big(1-\frac{1}{2}e^{\frac{b-m}{s}}\big)$$

$$=\frac{e^{\frac{b-m}{s}}}{2000}\big(1000-e^{\frac{m-n}{s}}-999e^{\frac{m}{s}}\big)$$

$$\geq\frac{e^{\frac{b-m}{s}}}{2000}\big(1000-e^{\frac{-1000m}{s}}-999e^{\frac{m}{s}}\big), \text{ since } n\geq 1001m$$

$$>0 \text{ for all } s>1,000,000m$$

2) for $0\leq b<m$, the difference between (3) and (2) is

$$1-\frac{1}{1000}\big(\frac{1}{2}e^{\frac{b-n}{s}}\big)-\frac{999}{1000}\big(1-\frac{1}{2}e^{-\frac{b}{s}}\big)-\big(1-\frac{1}{2}e^{\frac{b-m}{s}}\big)$$

$$=\frac{e^{\frac{b-m}{s}}}{2000}\big(1-e^{\frac{m-n}{s}}+999e^{\frac{m-2b}{s}}\big)$$

$$\geq\frac{e^{\frac{b-m}{s}}}{2000}\big(1-e^{\frac{-1000m}{s}}+999e^{\frac{-m}{s}}\big), \text{ since } n\geq 1001m \text{ and } b<m$$

$$>0 \text{ for all } s>m$$

3) for $m\leq b<n$, the difference between (3) and (2) is

$$1-\frac{1}{1000}\big(\frac{1}{2}e^{\frac{b-n}{s}}\big)-\frac{999}{1000}\big(1-\frac{1}{2}e^{\frac{b}{s}}\big)-\frac{1}{2}e^{\frac{m-b}{s}}$$

$$=\frac{1}{2000}\big(2+e^{-\frac{b}{s}}(999-1000e^{\frac{m}{s}}-e^{\frac{b-n}{s}}\big)$$

$$>\frac{1}{2000}\big(1+e^{-\frac{b}{s}}(999-1000e^{\frac{m}{s}})\big) \text{ since } b<n$$

$$> 0 \text{ for all real } s.$$

4) for $b \geq n$, the difference between (3) and (2) is

$$1 - \frac{1}{1000}\big(1 - \frac{1}{2}e^{\frac{n-b}{s}}\big) - \frac{999}{1000}\big(1 - \frac{1}{2}e^{\frac{b}{s}}\big) - \frac{1}{2}e^{\frac{m-b}{s}}$$

$$= \frac{e^{-\frac{b}{s}}}{2000}\big(999 + e^{\frac{n}{s}} - 1000e^{\frac{m}{s}}\big)$$

$$\geq \frac{e^{-\frac{b}{s}}}{2000}\big(999 + e^{\frac{1001m}{s}} - 1000e^{\frac{m}{s}}\big) \text{ since } n \geq 1001m$$

$$> 0 \text{ for all real} s.$$

Thus, (3) is strictly greater than (2) for all real b. Recall that this means that, for every possible outcome O_a of GPOF $\oplus Z$ and of $P_{\frac{1}{1000}} \oplus Z$, the latter has a strictly higher probability of an outcome at least as good as O_a. Therefore, by Stochastic Dominance, $P_{\frac{1}{1000}} \oplus Z$ is strictly better than GPOF $\oplus Z$. Then, by Prospect Separability, so too is $P_{\frac{1}{1000}}$ strictly better than GPOF. □

References

Adams, C. J.; Crary, A.; and Gruen, L., 2023. Future-oriented effective altruism: What's wrong with longtermism? In *The Good It Promises, the Harm It Does: Critical Essays on Effective Altruism* (Eds. C. J. Adams; A. Crary; and L. Gruen). Oxford University Press. (cited on page 2)

Arrhenius, G., 2000. An impossibility theorem for welfarist axiologies. *Economics and Philosophy*, 16 (2000), pp. 247–66. (cited on page 7)

Bader, R., n.d. Person-affecting population ethics. Unpublished manuscript. (cited on page 9)

Bader, R. M., 2022. Person-affecting utilitarianism. In *The Oxford Handbook of Population Ethics*, 251. Oxford University Press. (cited on pages 8 and 16)

Beckstead, N. and Thomas, T., 2021. A paradox for tiny probabilities and enormous values. *Noûs*, (2021). (cited on page 14)

Blackorby, C.; Bossert, W.; and Donaldson, D., 2005. *Population Issues in Social-Choice Theory, Welfare Economics and Ethics.* Cambridge University Press, New York. (cited on page 20)

Bramble, B., 2023. The heart of the problem with longtermism. Unpublished manuscript. (cited on page 2)

Broome, J., 2004. *Weighing Lives.* Blackwell. (cited on pages 9, 10, and 11)

Buchak, L., 2022. How should risk and ambiguity affect our charitable giving? Global Priorities Institute Working Paper Series. (cited on page 2)

Chappell, R. Y., 2023. X-risk agnosticism. Available at https://rychappell.substack.com/p/x-risk-agnosticism. (cited on page 1)

Cowen, T., 1992. Consequentialism implies a zero rate of intergenerational discount. In *Justice Between Age Groups and Generations*. Yale University Press. (cited on page 12)

Cowen, T. and Parfit, D., 1992. Against the social discount rate. In *Justice Between Age Groups and Generations (Philosophy, Politics, and Society)*, pp. 144–61. Yale University Press. (cited on page 1)

Crary, A., 2023. The toxic ideology of longtermism. *Radical Philosophy*, 214 (2023), pp. 49–57. (cited on page 2)

Cremer, C. Z. and Kemp, L., 2021. Democratising risk: In search of a methodology to study existential risk. *SSRN*, (2021). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3995225. (cited on page 2)

Greaves, H., 2016. Cluelessness. In *Proceedings of the Aristotelian Society*, vol. 116, 311–339. Oxford University Press. (cited on page 4)

Greaves, H. and MacAskill, W., 2021. The case for strong longtermism. Global Priorities Institute Working Paper Series. (cited on pages 1, 2, and 22)

Harsanyi, J. C., 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63, 4 (1955), pp. 309–21. (cited on page 11)

Henning, T., 2022. *Die Zukunft der Menschheit – soll es uns weiter geben?* J.B. Metzler. (cited on page 2)

Holtug, N., 2007. On giving priority to possible future people. In *Hommage á Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz*. (cited on page 8)

Huemer, M., 2008. In defence of repugnance. *Mind*, 117, 468 (2008), 899–933. (cited on page 10)

Leibniz, G. W., 2005. *Confessio Philosophi: Papers Concerning the Problem of Evil, 1671-1678*. Yale University Press. (cited on page 4)

Lenman, J., 2022. How effective altruism lost the plot. *iai news*, (2022). Available at https://iai.tv/articles/how-effective-altruism-lost-the-plot-auid-2284. (cited on page 2)

MacAskill, W., 2022. *What We Owe the Future*. Oneworld, London. (cited on pages 2 and 22)

MacAskill, W. and Mogensen, A., 2021. The paralysis argument. *Philosophers' Imprint*, 21, 15 (2021), pp. 1–17. (cited on page 4)

McMahan, J., 1981. Problems of population theory. *Ethics*, 92, 1 (1981), pp. 96–127. (cited on page 19)

Nebel, J., 2018. The good, the bad, and the transitivity of *better than*. *Noûs*, 52, 4 (2018), pp. 874–99. (cited on page 10)

Ord, T., 2020. *The Precipice: Existential Risk and the Future of Humanity*. Bloomsbury, London. (cited on pages 1 and 22)

Parfit, D., 1984. *Reasons and Persons.* Oxford University Press, Oxford. (cited on pages 1, 4, 6, 9, and 19)

Pettigrew, R., 2022. Effective altruism, risk, and human extinction. Global Priorities Institute Working Paper Series. (cited on page 2)

Pivato, M., 2014. Additive representation of separable preferences over infinite products. *Theory and Decision*, 77, 1 (2014), pp. 31–83. (cited on page 20)

Plant, M., 2023. Book review: William MacAskill, What We Owe The Future: A Million-Year View (One World Publications, London, 2022), pp. 246. *Utilitas*, (2023). (cited on pages 2 and 6)

Ramsey, F. P., 1928. A mathematical theory of saving. *The Economic Journal*, 38, 152 (1928), pp. 543–59. (cited on page 1)

Russell, J. S., 2023. On two arguments for fanaticism. *Noûs*, (2023). (cited on page 18)

Setiya, K., 2022. The new moral mathematics. *Boston Review*, (2022). Available at https://www.bostonreview.net/articles/the-new-moral-mathematics/. (cited on pages 2 and 6)

Sidgwick, H., 1907. *The Methods of Ethics, 7th edn.* Macmillan, London. (cited on pages 1 and 9)

Stock, K., 2022. Effective altruism is the new woke. *UnHerd*, (2022). Available at https://unherd.com/2022/09/effective-altruism-is-the-new-woke. (cited on page 2)

Tarsney, C., 2020. Exceeding expectations: stochastic dominance as a general decision theory. Global Priorities Institute Working Paper Series. (cited on page 29)

Tarsney, C. and Thomas, T., 2020. Non-additive axiologies in large worlds. Global Priorities Institute Working Paper Series. (cited on page 2)

Temkin, L., 2011. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning.* Oxford University Press. (cited on page 10)

Thomas, T., 2022a. The asymmetry, uncertainty, and the long term. *Philosophy and Phenomenological Research*, (2022). (cited on pages 2 and 20)

Thomas, T., 2022b. Separability and population ethics. *The Oxford Handbook of Population Ethics*, (2022), 271–296. (cited on pages 7, 20, and 28)

Thorstad, D., 2023a. High risk, low reward: A challenge to the astronomical value of existential risk mitigation. *Philosophy and Public Affairs*, (2023). (cited on pages 1 and 27)

Thorstad, D., 2023b. Three mistakes in the moral mathematics of existential risk. Global Priorities Institute Working Paper Series. (cited on pages 1 and 27)

Trammell, P., 2021. Patient philanthropy in an impatient world. Unpublished report, available at https://docs.google.com/document/d/1NcfTgZsqT9k30ngeQbappYyn-UO4vltjkm64n4or5r4/edit. (cited on page 3)

Van Liedekerke, L., 1995. Should utilitarians be cautious aboutan infinite future? *Australasian Journal of Philosophy*, 73, 3 (1995), pp. 405–7. (cited on page 9)

Walen, A., 2022. The atrocious conclusion: Why totalism-based longtermism is wrong. Unpublished manuscript. (cited on page 2)

Wilkinson, H., 2021. Infinite aggregation: Expanded addition. *Philosophical Studies*, 178 (2021), pp. 1917–49. (cited on page 9)

Wilkinson, H., n.d. Egyptology and fanaticism. Unpublished manuscript. (cited on page 18)

Wolf, E. and Toon, O., 2015. The evolution of habitable climates under the brightening sun. *Journal of Geophysical Research: Atmospheres*, 120, 12 (2015), 5775–5794. (cited on page 1)

Wolfendale, P., 2022. The weight of forever: Peter Wolfendale reviews What We Owe The Future. *The Philosopher*, November (2022). (cited on page 2)