The (Im)possibility of Prudence:
Population Ethics for Person-Stages

Marina Moreno

Abstract

This paper develops a largely neglected parallel between prudence and population ethics. Prudence is generally understood to be concerned with the balancing of well-being over time. How, *precisely*, well-being ought to be balanced over time, however, is a fervently debated question. I argue that developing a standard guiding such evaluations is exceedingly challenging. This is due to an often overlooked fact about prudence, namely that it shares a structural similarity with population ethics: In both contexts, we assess the comparative value of populations of person-stages/people, which may vary in number and level of well-being. Based on this analogy, I show that the development of an adequate theory of prudence runs into very similar impossibility results as obtain in population ethics. In particular, I prove that Arrhenius' fifth impossibility theorem can be applied to prudence. I develop and compare four possible answers to this challenge. First, I discuss the possibility of accepting the very repugnant conclusion intrapersonally. Second, I present and further develop Donald Bruckner's Minimax Regret approach, which gives up Transitivity. Third, I apply Jacob Nebel's Lexical Threshold View to prudence and critically evaluate it. Lastly, I introduce what I call the Negative Lexicality View, which is based on Lexical Threshold View but overcomes some of its problems.

Keywords Prudence, Population Ethics, Axiology, Diachronic Well-Being, Transitivity, Completeness, Very Repugnant Conclusion

1 Introduction

Prudence is generally understood to be concerned with the balancing of well-being over time. For instance, we deem it prudent to undergo a painful dental treatment if it helps us avoid greater pain in the future. Even though we thus sacrifice some well-being at the time of the treatment, the future avoidance of greater pain makes it a prudentially choice-worthy endeavour. How, precisely, well-being ought to be balanced over time, however, is a fervently debated question. While some authors have endorsed a maximisation principle over lifetime well-being (e.g. Sidgwick 1901, Price 2002), others have suggested averaging principles (e.g. Bricker 1980, Pettigrew 2019) and maximin principles (e.g. Bruckner 2003). This controversy has led as far as some authors (e.g., Huckfeldt 2011) suggesting that there is no universal prudential standard at all and that prudence is ultimately a matter of contingent personal commitments.

In this paper, I aim to show that matters are even worse than this. On the one hand, I shall argue contra Huckfeldt that the existence of a universal standard of prudence is both desirable and plausible. On the other hand, however, I argue that developing such a standard is exceedingly challenging. This is due to an overlooked structural fact about prudence, namely that it shares a crucial similarity with population ethics: In both contexts, we assess the comparative value of a population of person-stages/people, who may vary in number and level of well-being. I will show that the development of an adequate theory of prudence runs into very similar impossibility results as obtain in population ethics. I consider various possible answers to this challenge and discuss some general lessons to be drawn from it regarding well-being, prudence, and population ethics, as well as their conceptual connections.

The paper is organised as follows. In section 2, I examine the concept of prudence in greater detail and formally describe what I take to be one of the central issues of prudence, namely defining a betterness relation over sequences of person-stages with varying levels of well-being. In section 3, I will illuminate the close structural parallels between the issue of prudence, as explained in section 2, and population ethics. In particular, I shall show that Arrhenius's fifth impossibility theorem (cf. Arrhenius 2000, Arrhenius 2003, Arrhenius 2011) can be applied to prudence, arguably posing an even greater challenge for prudence than it does for population ethics. Section 4 examines potential escape routes by examining various relaxations of desiderata for principles of prudence. Finally, section 5 extracts key insights from the preceding investigation that hold implications for prudence and well-being and possibly even reverberate back to population ethics.

2 Prudence as Balancing Well-Being Over Time

Suppose you are contemplating whether to schedule a dental appointment. You are aware of the potential discomfort the treatment may bring and the opportunity costs accrued by committing time to the dentist's appointment. However, you also understand that undergoing this treatment will preclude more severe pain and lost time in the future. Intuitively, the prudent choice in such a situation would be to proceed with the filling. Indeed, such a decision appears to be a paradigmatic case of prudence. We can formalise this scenario with the following matrix:

Going to the Dentist

	t_1	t_2
w_1	-5	10
w_2	10	-50

In the first world w_1, you undertake the painful treatment at time t_1, resulting in a healthier condition at t_2. In the alternate world, w_2, you opt against the treatment at t_1, but endure exacerbated pain at t_2. These comparative gains and losses are formalised using integers. Although I refrain from endorsing whether a set of rational values and preferences are definitively and precisely representable by a single utility function, it serves as a practical conceptual tool to formalise and compare levels of well-being (to the extent that such comparisons are comprehensible).

Based on this formalisation, we can elucidate your reasoning when deciding whether to proceed with the dental treatment as follows: If you decide to see the dentist, you find yourself in w_1, meaning that your well-being level stands at -5 at t_1, but is compensated by a positive well-being level of +10 at t_2. In contrast, if you refrain from seeing the dentist, your well-being level will be comparatively higher at t_1, i.e., +10, but this is counterweighed by a significantly lower well-being level of -50 at t_2. Given some plausible method of aggregating the well-being levels of each world, w_1 appears to be the overall superior choice for you, thus making it the prudent selection. More generally, then, prudence appears to offer guidance on balancing well-being levels over time, i.e. balancing the well-being levels of different *person-stages* over time. That is, a theory of prudence provides insights on evaluating which sequences of person-stages, characterised by their level of well-being, are better or more choice-worthy from a self-regarding perspective.

We can distil at least three preconditions that must be satisfied for the above reason-

ing to be intelligible. Firstly, to comprehend the balancing of well-being levels of different person-stages, we must be capable of meaningfully *separating* those person-stages. Assuming such separability does not necessarily commit us to any metaphysically demanding notion of separated person-stages. We must merely be able to separate the levels of well-being, i.e., meaningfully attribute separate levels of well-being at separate times. The person-stages can then be characterised as the period of time in which the person had the respective level of well-being. Person-stage then simply refers to *the person in the time period with the respective well-being level*. Secondly, these well-being levels must be, in one form or another, comparable. If each separable level of well-being were entirely incomparable, rendering the evaluations of more and less desirable well-being states meaningless, then the question regarding which levels are prudentially preferable appears moot. Lastly, we must be capable of forming judgements regarding their aggregations. That is, we should not only be able to meaningfully distinguish, for example, the distinct well-being levels at different times based on whether or not we receive the treatment, but also appraise the prudential value of the overall sequences of person-stages in a specific world, for instance, how prudentially worthy the sequences -5, 10 and 10, -50 are. In what follows, I shall call such sequences of person-stages at a specific world *lives*.

Based on these conditions, we can now capture (one of) the problem(s) of prudence as follows:

Problem of Prudence: Find an axiology of lives, i.e., a betterness or choice-worthiness relation over the set of all possible lives, where

- a **life** is an ordered and finite, but arbitrarily large, sequence of time-indexed person-stages;

- a **person-stage** is a finite time-indexed period of time in a person's life that is characterised by the respective level of well-being during that period (and not specified any further).

The present paper is concerned with the problem of prudence. For example, a summative principle, which posits that the net total well-being determines the overall and comparative value of lives, can be considered a potential candidate for such a principle of prudence. Whether there even is an ultimately satisfactory principle, however, is partially what I aim to explore in this paper.

4

Before moving on, let me add some final remarks about my understanding of prudence. First, given that various authors have identified the basic structure of the question concerning how to balance well-being over time (e.g., Bruckner 2003, 33; Huckfeldt 2011, 43; Brink 2003, 215; Dorsey 2018, 2901: Bricker 1980, 381) with differing philosophical commitments, encompassing various views on well-being, the issues I aim to investigate appear to be at least somewhat neutral regarding the nature of well-being. Crucially, merely endorsing the relatively minimal conditions of separable and comparable levels of well-being which are, in some form, internally aggregable, is sufficient to formalise and scrutinise the issue of balancing well-being over time in the way I proposed. These conditions seem compatible with all three common traditional approaches to well-being, i.e., hedonism, attitude-based views, and objective list views (cf. Parfit 1984, app. I). I will revisit the question of how various theories of well-being and the potential rejection of any of the three conditions might influence the choice and development of theories for prudence. But at least prima facie, my understanding of prudence seems broadly applicable.[1]

Second, there are both self-regarding as well as moral reasons to think that finding a universal principle of prudence that provides an axiology of lives is important. Some authors, such as Huckfeldt, have recently argued that prudential principles should not be understood as universally valid, but rather as based on contingent commitments. I hold, contra Huckfeldt, that a principle of prudence must be at least somewhat universal. This is due to the fact that at least *some* prudential evaluations hold independently of subjective evaluations. Consider Huckfeldt's argument, who bolsters his claim by emphasizing that it is *rationally* permissible to neglect the considerations of our future person-stages, suggesting that temporal extension does not obligate us to regard our future person-stages as relevant (Huckfeldt 2011, 44f). However, even if Huckfeldt is

---

1. The following notes might provide a hint of my views on the compatibility of theories of well-being and prudence as sequence-comparability. Firstly, most hedonists consider conscious experience to be separable and comparable over time, and they find the aggregation of their value to be meaningful (e.g. Sidgwick 1901, Bentham 1780). Similarly, most proponents of attitude-based views advocate for some form of comparability and aggregation over time (e.g. Dorsey 2021, ch. 9). While many such advocates would likely also support the claim that these levels of well-being are meaningfully separate, it at the very least plausibly holds if our attitudes change over time, such that there are discrete levels of attitude-satisfaction that can be divided into before and after such a transformation. Indeed, it is this separability that gives rise to the kinds of puzzles that, e.g., Paul (2014) and Pettigrew (2019) grapple with. Pettigrew indeed likewise formalises the problem as a question of how to *aggregate* the distinct evaluations of distinct selves. Finally, provided that objects constituting well-being can be separated in time, and provided that the value of these objects can be compared and aggregated, objective list theories also seem to meet all three conditions. Separability, in particular, seems to hold with respect to *some* objects claimed to be valuable by objective list theorists, even if other objects might be valuable in an atemporal and non-separable manner.

correct in asserting that prudential concerns are not necessarily tantamount to require-
ments of *rationality*, this does not mean that the prudential concerns themselves are
wholly subjective. Consider for instance a person who, at t_0, subjectively prefers for all
of their own future person-stages to have negative well-being. Even if such preferences
are not excluded by rationality (which is itself controversial of course), as Huckfeldt
proposes, it seems abundantly clear that the person's life, as a subjective unit persisting
through time, will be *better* for them if they do not make their future person-stages suffer.
This evaluative assertion seems to remain valid regardless of the individual subjective
preference, rendering it universal rather than subjective.

Lastly, finding a principle of prudence that enables us to discern which lives are
better for an individual is also paramount from the perspective of moral theory: The
ability to determine which decisions benefit others constitutes a fundamental component
of any plausible moral theory. Such an aptitude presumes the capability to assess what
is better or worse for others, encompassing the evaluation of sequences of person-stages
over time. For instance, if I were to decide between minimising or maximising well-
being of future person-stages on someone else's behalf, opting for maximising the well-
being would clearly be the morally correct choice. Hence, if we perceive prudence as an
evaluative theory concerning the goodness (or choice-worthiness) of lives, any credible
moral theory must integrate a plausible theory of prudence.

Note, however, that the plausibility of the existence of *some* universal constraints
on prudence does not imply that there must be a fully characterised axiology of lives.
Whether this is possible and plausible is partially what I explore in this paper.

3 Prudence and Population Ethics

The traditional stance on prudence is a temporally neutral additive maximizing princi-
ple, championed by Henry Sidgwick (1901) and expanded upon by Bruce Price (2002).
However, other principles have also gained traction, such as principles maximizing aver-
age well-being (Bricker 1980; Pettigrew 2019), and various minimax well-being principles
(Bruckner 2003). Interestingly, most of the authors attempting to solve the problem of
prudence neglect an important dimension when considering live comparisons, namely
that lives can be unequally long. This oversight is surprising, given that neglecting this
important aspect trivialises prudence. For instance, maximising and averaging princi-
ples converge under an assumption of equally long lives.[2] One aim of this paper is to

---

2. An explicit exception is Bricker, who argues against maximising overall well-being and for an
averaging principle by appealing to cases of unequally long lives (Bricker 1980, 392). However, in doing

remedy this oversight. In order to provide an answer to the problem of prudence, then, it is important to find desiderata for a principle defining a betterness relation over *all possible* lives, rather than merely equally long possible lives.

The complication of unequally large sets to be compared is familiar from a different domain of practical philosophy, namely population ethics. Indeed, I argue that it this is not the only shared feature, but that population ethics bears a striking structural similarity to prudence more broadly, such that many of the desiderata for a theory for population ethics can be straightforwardly applied to prudence as well.

Popularised by Parfit, the crux of population ethics asks: Given various populations of people with varying levels of well-being (which could range from egregiously low to extraordinarily high), how do we determine which ones are better than others, or, deontically put, which ones are more worth creating than others? Importantly, the populations compared in population ethics are usually fully characterised by the number of people in the respective population as well as their individual life-time well-being. Population ethics, as it is ordinarily understood, is thus importantly about population well-being. While other moral considerations in creating populations are becoming more frequently discussed as well, the original problem is posed in terms of well-being, i.e., in terms of which populations are better or worse *with respect to the population well-being.*

Though often overlooked,[3] I believe the structural similarity between population ethics and prudence as I sketched it in section 2 is evident. Just as in population ethics the people in the populations are fully characterised in terms of their well-being, the same holds in the context of prudence for the individual person-stages. Similarly, the number of people in a population is usually understood to be finite, though arbitrarily large, just as sequences of person-stages are usually understood to be finite, though arbitrarily long. The betterness relation for population ethics is usually defined in terms of some form of aggregation of the population, just as the betterness relation for prudence is usually defined in terms of some form of aggregation of the person-stages. Overall, then, we can put the problem of prudence analogously to the problem of population ethics: Given different sequences of person-stages with varying levels of well-being (potentially

---

so he neglects that his own averaging principle is vulnerable to similar objections as he presents to the maximisation view. In particular, his view fails to meet Dominance Addition, as defined in section 4.

3. A notable exception is Nebel (2019), who develops a different parallel in detail, namely between population ethics and intrapersonal cases of being brought into existence, where in the latter case it is probabilities that are aggregated, rather than person-stages. He also notes in passing that a unified solution to prudence and population ethics may be desirable in Nebel 2022, 203. Other authors have considered the analogy between populations and individuals in other contexts, such as Broome 2017 or Hurley 1989. Broome considers how a disuniting personhood matephysics supports the utilitarian principle, while Hurley explores the analogy between social choice and the aggregation of individual reasons.

from extremely low to exceptionally high), how do we determine which ones are better? To put it concretely, by replacing the individual persons in population ethics with individual person-stages and introducing a temporal ordering to these person-stages, we can transform any question in population ethics into a corresponding question in prudence.

While I believe the similarity to be striking indeed, there are also some important disanalogies that need to constrain, or at least inform, the application of arguments in population ethics to prudence. Firstly, populations of people are usually not (temporally) ordered, while sequences of person-stages are. Some authors, such as Velleman (1991), argue that the temporal structure of someone's life can have a certain kind of value or disvalue above and beyond the mere sum of its temporally located well-being parts. Velleman argues for this conclusion by pointing to different kinds of life-arcs, e.g., whether our lives get continuously better or continuously worse, arguing that some arcs are better than others. While I will not further examine Velleman's proposal in this paper, partly because the problems I shall outline cannot be solved by appealing to life-arcs either way, it is important to note that similar temporal considerations do not usually play any role in population ethics, making for a possibly important disanalogy.[4]

While I do not find myself convinced that life arcs matter in this way, I argue that the present investigations should still be interesting for proponents of this view for the following reason: While life arcs might be *one* relevant source of the value of a life, it is likely not solely determinate. For instance, a life that starts at a well-being level of 1 and consistently increases is likely worse than a a life that starts at a well-being level of 10 and consistently increases, even if their arc is functionally the same, or very similar. The crucial difference here does not seem to lie in the arc, but rather in some aggregation of the well-being levels. The present investigations can then be viewed as determining this underlying value, which may be supplemented by the value that depends on the arc.

Secondly, most views of persons and person-stages regard the two as relevantly different in terms of their separateness. That is, while separate person-stages are usually still understood to be part of the same subject, and often also the same moral patient, different people are generally regarded as distinct subjects and distinct moral patients. This may lead to important disanalogies. For instance, Rawls famously argued that the "separateness of persons" has significant moral import (Rawls 1971, 23, 163), i.e., that the fact that distinct people have fundamentally distinct subjective loci must be taken seriously by any adequate moral theory. There are, of course, theories of personal iden-

---

4. This is because in population ethics, an Anonymity condition, i.e., evaluative invariance under identity permutations, is usually implicitly assumed. It may be interesting, however, to explore population ethics without such a condition, which would result in an even tighter analogy between prudence and population ethics.

tity that posit a similar metaphysical separateness between persons and person-stages (e.g. Parfit 1984). If we held that a genuinely new person comes into existence at every moment in time, prudence would be even more analogous to interpersonal population ethics, as we would now genuinely be aggregating distinct persons even in the intrapersonal case. For such a view, then, the analogies I will draw in what follows will apply even more closely. For the present purposes, however I shall work with a more common sense understanding of personhood and personal identity, according to which people generally (notwithstanding some extreme cases) stay the same person within their lives.

Given this difference between persons and person-stages, we note at least two important conceptual differences: Firstly, the aggregation of person-stages can relevantly be understood as *compensation*, since the very same person experiences each of the stages. In the interpersonal case, however, axiological aggregation across persons cannot be understood as compensation and is often problematised (e.g. Taurek 1977). Secondly, most so-called person-affecting views in population ethics are based on the intuition that while we have a moral reason to make existing people happy, the mere fact that a person's life will be happy does not give us any moral reason to create them. This person-affecting feature, i.e., whether our choice will actually affect an existing person, cannot be directly translated to prudence: Prudential issues standardly always affect some person. While some population ethics theories may thus find their straightforward prudential counterparts (such as totalism), person-affecting theories are likely not as easily translated into what we might want to call "person-stage-affecting" theories.

This, too, does not undermine the relevance of the analogy. Firstly, the analogy is still entirely intact for a Parfitian-type picture of personal identity, which suggests that there is a similar separateness between person-stages as between persons (cf. Parfit 1984). Secondly, even if we deny this picture, population ethics theories inspired by person-affecting intuitions may still be translatable to prudence if adjusted by the right tweaks. In fact, section 4.2 presents a theory of prudence that is partly inspired by Meacham's person-affecting view in population ethics (2012). Lastly, and most importantly, the literature on population ethics that is *not* based on person-affecting intuitions is still vast, such that there is still ample space for a fruitful application.

Despite some disanalogies, then, I believe the structural similarity between prudence and population ethics to be significant and fruitful, as the following sections shall show. It is thus surprising that it does not seem to have received its due attention. This paper aims remedy this oversight.

The structural similarity between prudence and population ethics is fortunate insofar as it gives us a whole new pool of inspiration. In line with this, I will take seven

desiderata for an axiology of populations and translate them to the prudential context. I shall argue that they are at least as plausible to hold for prudence, if not more so. Yet, the similarity with population ethics will also turn out to be unfortunate in a different respect. As Gustaf Arrhenius famously demonstrated in a series of papers, population ethics is fraught with impossibility theorems (2000, 2003, 2009, 2011). The following section will demonstrate that prudence faces similar challenges.

3.1   Seven Desiderata

While I will keep the formalisms to a minimum in the main text, it is useful to set up the basic constituents of my formal framework at this point. Formally, I take a life to be a finite sequence of integers. The sequence index represents the time index, and the integer represents the level of well-being of the respective person-stage at this time. A prudential axiology of lives is a *binary betterness relation* over *all possible* lives. That is, we do not restrict the domain of lives, e.g., to equally long lives. The betterness relation is, as usual, built from the basic relation ≥ which is to be understood as "at least as good as". For any two lives L_x and L_y, it holds that if $L_x \geq L_y$ and $L_y \geq L_x$, then $L_x \sim L_y$, which is to be understood as "L_x is equally good as L_y". Furthermore, for any two lives L_x and L_y, it holds that if $L_x \geq L_y$ and $L_y \ngeq L_x$, then $L_x > L_y$, which is to be understood as "L_x is better than L_y". The present section examines seven desiderata that such an axiology of lives plausibly ought to meet.

The first two desiderata for an axiology of lives are structural. In order for a binary betterness relation to give us a genuine axiology, the respective relation should ideally generates a complete ordering. This leads to the following two desiderata:

Completeness: For any two lives L_x and L_y, $L_x \geq L_y$ or $L_y \geq L_x$ (or both).

Transitivity: For any three (commensurable) lives L_x, L_y and L_z, if $L_x \geq L_y$, and $L_y \geq L_z$, then $L_x \geq L_z$.

Having a complete ordering is both conceptually and practically desirable. While there may ultimately be reason to conclude that some lives are genuinely not comparable, this would require a significant underlying justification. This is especially the case because only a complete ordering guarantees us guidance in every possible situation. Similarly, transitivity is likewise conceptually and practically desirable. Betterness itself is generally understood to be a transitive concept, since otherwise resulting betterness-cycles seem to run counter the intuitive understanding of what it means that something is

better than something else. Additionally, such cycles leave agents vulnerable to money-pumps and can lead to inconsistent guidance in choices between multiple lives. While, again, there may be good reason to give up transitivity, doing so generates substantive theoretical and practical costs.

The next two desiderata, Egalitarian Dominance and Dominance Addition owe their names to analogous axioms in population ethics and can be grouped together as Dominance Principles.

> **Egalitarian Dominance**: Take any two lives L_x and L_y that are fully characterised by the following properties.
>
> - L_x and L_y are equally long.
> - Within both lives the well-being consistently stays the same.
> - The well-being in L_x is consistently higher than the well-being in L_y.
>
> If the above properties fully characterise L_x and L_y, then L_x is better than L_y.

More informally, Egalitarian Dominance holds that a life that is equally long but consistently and equally better than another ought to be judged as overall better. The matrix below illustrates this idea: L_1 must be better than L_2, given that L_1 has a consistently high well-being and L_2 has a consistently low well-being. If a theory of prudence fails to satisfy even Egalitarian Dominance, it seems to me, then the concepts of 'better' or 'worse', for a temporally extended person lose their meaning altogether.

Egalitarian Dominance: $L_1 > L_2$

	t_1	t_2	t_3	t_4	t_5
L_1	100	100	100	100	100
L_2	1	1	1	1	1

A similarly strong desideratum for a theory of prudence is the following.

Dominance Addition: Take any two lives L_x and L_y that are fully characterised by the following properties.

- The same number of person-stages as exist in L_x exist in L_y with a higher level of well-being.

- Additional person-stages with positive well-being exist in L_x.

If the above properties fully characterise L_x and L_y, then L_x is at least as good as L_y.

The intuitive idea of Dominance Addition is that adding positive well-being ought not make the respective life worse. Consider the following example of this idea.

Dominance Addition: $L_2 \succeq L_1$

	t_1	t_2	$\ldots$	t_n
L_1	90	–	–	–
L_2	100	5	$\ldots$	5

In the above case, Dominance Addition holds that L_2 ought not be strictly worse than L_1. That is, it cannot be the case that it is strictly the prudentially right choice to choose L_1. This, again, seems intuitively right. Increasing our counterfactual well-being and adding more positive well-being should at the very least be prudentially allowed.[5]

The next two desiderata, General Non-Elitism and General Non-Extreme Priority, are likewise borrowed from population ethics. In the prudential context, they can be viewed as desiderata of non-myopia: They both hold that small well-being improvements in one life should not generally outweigh possibly large well-being improvements for possibly many other person-stages in the other. Insofar as prudence is precisely concerned with balancing well-being over time, rather than deciding myopically for any at any particular time, they both seem to be very plausible constraints for prudence.

---

5. Note that this would also hold if in L_2 at t_1 there was a well-being of 5, and the well-being of 100 came later in L_2. The above example is merely the most obvious illustration of the condition, but the ordering is not relevant for it. This applies equally to the the conditions that will follow.

General Non-Elitism: Take any two lives L_x and L_y that are fully characterised by the following properties.

- L_x and L_y are equally long.

- There is some number (possibly 0) of person-stages in L_x and L_y with equal levels of well-being, i.e., they share a subpopulation of person-stages.

- There is one person-stage in L_x whose level of well-being, w_x, is exactly one unit higher than the well-being of one corresponding person-stage in L_y.

- There is some number n of person-stages in L_y with a well-being level w_x, where w_x is higher by at least more than one unit than the well-being of the same number n of person-stages in L_x.

If the above properties fully characterise L_x and L_y, then for any w_x there is some number n such that L_y is at least as good as L_x.

The intuition behind General Non-Elitism is the following: Making small comparative improvements at one time, should not prudentially outweigh comparatively (possibly significantly) larger well-being gains at other times for potentially many other person-stages. The following matrix illustrates an example of General Non-Elitism.

General Non-Elitism: $L_2 \succeq L_1$

	t_1	t_2	t_3	t_4
L_1	100	100	1	1
L_2	100	99	99	99

The fact that at t_2, I am slightly better off if I choose L_1 ought not entirely outweigh the large well-being gains I receive from choosing L_2 at t_3 and t_4. As mentioned above, this seems to be a plausible constraint on prudence; arbitrarily holding the marginal gain of one specific person-stage as determining overall prudential goodness seems very counterintuitive. If we were to choose prudentially on someone else's behalf, we might want to say that it must at the very least be morally allowed to choose L_2 for them, if it is not altogether morally required.

General Non-Extreme Priority builds on a very similar intuition as General Non-Elitism. Its formal definition is as follows.

General Non-Extreme Priority: Take any two lives L_x and L_y that are fully characterised by the following properties.

- L_x and L_y are equally long.

- There is some number (possibly 0) of person-stages in L_x and L_y with equal level of well-being, i.e., they share a subpopulation of person-stages.

- There is one person-stage in L_x whose level of well-being, w_x, is exactly one unit higher than the well-being of one corresponding person-stage in L_y.

- There is some additional number n of person-stages in L_y whose level of well-being is among the very best possible, while the same number n of additional person-stages in L_x have a well-being level barely above neutral (which we operationalise formally as a number between 1 and 3).

If the above properties fully characterise L_x and L_y, there is some number n such that L_y is at least as good as L_x.

The intuitive idea is as follows: Very small improvements at certain times cannot take categorical priority over raising the well-being of possibly many person-stages from very low positive levels to very high positive levels. The following matrix illustrates an example.

General Non-Extreme Priority: $L_2 \succeq L_1$

	t_1	t_2	t_3	t_4	t_5
L_1	100	100	-2	1	1
L_2	100	100	-3	1000	1000

Just because the person-stage at t_3 in L_1 is slightly better off than its counterpart in L_2, this should not take priority over the large well-being gains there are when comparing the person-stages at t_4 and t_5.

Given the conceptual similarity to General Non-Elitism, similar reasons speak in favour of General Non-Extreme Priority as a condition for prudence. Arbitrarily holding the marginal gain of one specific person-stage as determining overall prudential goodness seems very counterintuitive. And again, considering the case from a moral perspective

is illuminating: Choosing L_1 for someone else would not seem to adequately take their interest into account from a moral perspective.

Lastly, a good theory of prudence should be able to avoid intrapersonal analogues of the very repugnant conclusion[6]. The formal definition of this desideratum is as follows.

> **Non-Repugnance**: Take any two lives L_x and L_y that are fully characterised by the following properties.
>
> - L_x consists of some number of person-stages n_x with a well-being level that is among the very best possible levels.
>
> - There is some number n_{y1} person-stages in L_y with a well-being level that is (possibly significantly) negative.
>
> - There is an additional number n_{y2} of person-stages in L_y whose well-being level is 3, i.e., barely above 0.
>
> If the above properties fully characterise L_x and L_y, then it is not the case that for any n_x and n_{y1}, there is an n_{y2} such that L_y is better than L_x.

The intuition behind non-repugnance is the following: A long extended life that at times involves potentially significant suffering while at best being barely worth living should not generally be prudentially better than a shorter life with extremely high levels of well-being. Consider the following matrix as an example.

Non-Repugnance: $L_1 \succeq L_2$

	t_1	t_2	t_3	t_4	$\ldots$	t_n
L_1	1000	1000	1000	–	–	–
L_2	-1000	-1000	-1000	3	$\ldots$	3

An objection to Non-Repugnance might be raised at this point: the interpersonal case may be sufficiently dissimilar that the repugnance does not carry over to the intrapersonal case. Indeed, one could argue that in the interpersonal case, no single being experiences the many instances of low well-being, while things differ in the intrapersonal

---

6. I will not be concerned with the repugnant conclusion here, according to which a sufficient number of lives barely worth living can outweigh a smaller number of lives that are extremely good. Though similar, the very repugnant conclusion is generally understood to constitutes a bigger challenge than the repugnant conclusion, and thus the more important desideratum. For this reason I restrict my discussion to the very repugnant conclusion. 'Non-Repugnance' henceforth always refers the avoidance of the *very* repugnant conclusion.

case. In L_2, *someone* gets to live a very long life, albeit with a low level of well-being. I will return to the question of how repugnant the intrapersonal very repugnant conclusion really is. For now it suffices to note that, prima facie, it is plausible that L_1 should at least be prudentially *allowed*. A theory of prudence that tells us that we *must* go through agonising torture in order to secure a long life barely worth living, thereby foregoing a shorter but extremely happy life, seems to go wrong.

3.2   The Impossibility of Prudence

Having taken a first pass at defining desiderata for an adequate principle of prudence, we must now ask the following question: Which principles meet these desiderata? Unfortunately, the answer is none.

> **Theorem**: There is no betterness relation which jointly satisfies Transitivity, Egalitarian Dominance, Dominance Addition, General Non-Elitism, General Non-Extreme Priority and Non-Repugnance.[7]

I prove the theorem in the appendix; it is analogous to the proof for Arrhenius' fifth impossibility theorem. Given this theorem, it becomes apparent that prudence does not only bear some superficial structural similarity to population ethics. It also inherits (at least some of) its substantive problems.[8] These problems might indeed be *even more* severe for prudence for the following reason.

The impossibility theorems show that there is no complete axiological ordering satisfying Egalitarian Dominance, Dominance Addition, General Non-Elitism, General Non-Extreme Priority and Non-Repugnance. As a matter of axiology, this is problematic both for population ethics as well as prudence. However, since population ethics crucially concerns moral questions regarding other people, the link between axiology and deontic recommendations need not be as tight as it likely is for prudence. Popular theories of population ethics are sometimes criticized for not sufficiently taking other deontological constraints into account when evaluating the deontic recommendations implied by an underlying axiology. While finding a population axiology remains exceedingly challenging, this may not be so problematic if it is ultimately other considerations

---

7. Note that Completeness is not a part of the set that is not jointly satisfiable. I outlined it above because it will become important in the discussion that follows, as it will be given up by various escape routes.

8. Note, for instance, that there is also a prudential version of Arrhenius' sixth impossibility theorem if we replace Dominance Addition with a desideratum of Weak Non-Sadism, i.e., Non-Masochism for prudence.

that determine our ultimate deontic duties anyway. However, similar critiques are likely misplaced for prudence. For it is more plausible that what we ought to choose from a purely prudential perspective is simply what is better for us. In this sense, then, the impossibility of an axiology for individual lives might be even more problematic than the impossibility of an axiology for populations.

As argued, finding an adequate principle of prudence is important for various reasons. However, insofar as an adequate principle is (at least) characterised by the above desiderata, it is impossible to find one. We are thus tasked with finding escape routes out of the impossibility. This is what the rest of the paper is concerned with.

4 Escape Routes

The seven desiderata can be grouped into types of desiderata. First, Transitivity and Completeness are *structural principles*. They tell us what the axiological relation itself should look like. Their plausibility stems from intuitions regarding betterness and conditions for adequate action guidance.

Second, Egalitarian Dominance and Dominance Addition are *dominance principles*. Their plausibility stems from dominance considerations: For both desiderata, one life dominates the other in some sense. For Egalitarian Dominance this is straightforward: One life consistently outperforms the other. For Dominance Addition, this reasoning is extended to variable sequence orderings and unequally long lives. But the basic spirit is the same: One life outperforms the other in all the relevant respects.

Third, General Non-Elitism (GNE), General Non-Extreme Priority (GNEP), and Non-Repugnance are *principles of well-being trade-offs*. What makes General Non-Elitism and General Non-Extreme Priority plausible is that small improvements for a single person-stage should not outweigh possibly large improvements for possibly many other person-stages. In this sense, they are also expressing the non-myopic spirit of prudence: Prudential decisions should not exclusively be made from the perspective of a single person-stage, and in this sense they should not be myopic. In this respect, they are also *non-myopia principles*. But their non-myopia is specifically about well-being trade-offs, since *small* improvements should not generally *outweigh larger* improvements. Similarly, Non-Repugnance is about the well-being trade-offs between low positive well-being and high negative/positive well-being: We do not want it to be prudentially required to accept strongly negative well-being instead of strongly positive well-being, simply to attain a very long life of very low positive well-being.

With these observations, we can note the following. Escaping the impossibility theo-

rem by giving up on either of the dominance principles is not promising. When one life outperforms the other in essentially every relevant respect, a principle of prudence should respect this judgement. The following sections thus examine escape routes based on the other four desiderata. That is, I will first examine giving up Non-Repugnance via adopting Totalism. Then I will examine giving up Transitivity via adopting Minimax Regret (Bruckner 2003) or giving up GNE via adopting a Lexical Threshold View (Nebel 2022), which will end up amounting to quite similar proposals in a structural sense. Lastly, I consider giving up GNEP by introducing what I call the Negative Lexicality View.

4.1 Giving up Non-Repugnance

So-called totalism gives up Non-Repugnance but meets the rest of the desiderata. It can be formally stated as follows:

> L_x is at least as good as L_y iff $T_{L_x} \geq T_{L_y}$, where T_{L_i} refers to the total sum of well-being in L_i.

It is easy to verify that this principle meets all the other desiderata. So what might speak in favour of giving up Non-Repugnance? Recall the objection we raised against Non-Repugnance. While the very repugnant conclusion is generally rejected in population ethics, the intrapersonal case might be relevantly different. This is due to the following disanalogy between population ethics and prudence: Most views of persons and person-stages regard the two as relevantly different in terms of their separateness. That is, while separate person-stages are usually still understood to be part of the same subject, and often also the same moral patient, different people are generally regarded as distinct subjects and distinct moral patients. As noted in section 3, this may lead to important disanalogies. For instance, Rawls famously argued that the "separateness of persons" has significant moral import (Rawls 1971, 23, 163), i.e., that the fact that distinct people have fundamentally distinct subjective loci must be taken seriously by any adequate moral theory. There are of course theories of personal identity that posit a similar metaphysical separateness between persons and person-stages (e.g. Parfit 1984). If we held that a genuinely new person comes into existence at every moment in time, prudence would be even more analogous to interpersonal population ethics, as we would now genuinely be aggregating distinct persons even in the intrapersonal case. For such a view, then, the analogies I draw apply even more closely. For the present purposes, however, I am working with a more common sense understanding of personhood and personal identity, according to which people generally (notwithstanding some extreme cases) stay the same person within their lives.

Indeed, Brink argues that it is precisely the fact that it is the *very same person* who can be compensated by later person-stages for current person-stage's sacrifices is the distinctive feature of prudence (Brink 2003, 223).[9] We thus noted that in the interpersonal case, no single being experiences the many instances of low well-being, while things are different in the intrapersonal case. Consider again our example above:

Non-Repugnance: $L_1 \geq L_2$

	t_1	t_2	t_3	t_4	$\ldots$	t_n
L_1	1000	1000	1000	–	–	–
L_2	-1000	-1000	-1000	3	$\ldots$	3

In L_2, *someone* gets to live a very long life, albeit with a low level of well-being. We find the very repugnant conclusion very repugnant for the interpersonal case *because* no one actually benefits from the large population with low levels of well-being. But in the intrapersonal case, someone *does* benefit. Therefore, the intrapersonal version of the very repugnant conclusion is not actually repugnant, or so the argument might go.

Above I already noted one possible counterargument. Even if we would want to prudentially *allow* for people to choose extreme suffering in order to secure the long extended life, it seems odd to have it be a prudential *requirement*. However, there is a further argument bolstering Non-Repugnance as a criterion.

The intuition leading us to hold that L_2 could be better than L_1 might at least partly be caused by the fact that a long life is itself valuable to us, such that we intuitively interpret L_2 falsely, i.e., intuitively attribute a higher level of well-being to the individual person-stages because we imagine them being better off given their long life.[10] To eliminate this confounder, consider the following interpretation of the above matrix: Each period of time lasts for one million years. In L_1 you exist for three million years at a very high level of well-being, your life is consistently filled with the best possible things. In L_2, you exist for n million years, but the first three million are filled with agonising torture and suffering, while the rest of the time you are at a level of well-being that is just barely worth living. Put this way, both lives end up being very long, such that

---

9. This idea of intrapersonal compensation even seems to lie at the heart of the concept of well-being. For instance, Scanlon argues against a distinct concept of well-being as such by denying that there is such a "sphere of compensation" at all (Scanlon 1998, 127).

10. Note that this error theory of our intuitions does not depend on a value-based theory of well-being. We might be led astray either because we subconsciously attribute a better hedonic state, a higher value-satisfaction or a life with an additional valuable "object" to a person with a sufficiently long life.

our intuitions likely attribute much less value to the mere fact of temporal extension. I at least find my intuitions become clear at this point. It also seems intrapersonally extremely repugnant to judge L_2 to be better than L_1. This seems especially clear when considering the moral perspective again: If we had to make a decision on behalf of someone else, it would seem outright cruel to choose L_2 for them.

This error theory may not successfully solve the problem, of course. My proposed reinterpretation may not actually move our intuitions sufficiently. One may also reject fanciful cases with such long lives outright, and argue that our intuitions are likely to be led astray *even more* if the time horizons are longer than anything we have ever experienced. If either of these objections are sound, we may want to opt to accept the intrapersonal very repugnant conclusion after all, which allows us to escape the impossibility theorem by giving up on Non-Repugnance. The remaining desiderata could then be satisfied, for instance, by a simple maximising principle, according to which lives are ordered based on their net sum of well-being. Taking this route would thus vindicate the traditional stance with a renewed sense of the associated theoretical costs.

4.2 Giving up Transitivity

Given the dominance (Egalitarian Dominance and Dominance Addition) principles, it is the *iterative* application of General Non-Extreme Priority and General Non-Elitism which leads via Transitivity straight to the very repugnant conclusion. That is, General Non-Extreme Priority and General Non-Elitism enable us to construct a series of lives by consistently altering lives step by step, such that each individual alteration makes the lives better according to General Non-Extreme Priority or General Non-Elitism. But by the end of this procedure we are comparing two lives with big well-being differences, such that an overall transitive judgement between the first and the last life considered does no longer seem right (cf. appendix (step 3 and 4 of the proof) to see how the theorem is based on such a procedure). Together, these series lead us to the very repugnant conclusion. At least one of the series constructed in this way may actually itself already be viewed as problematic. Consider the following construction of a series of lives, where L_1 is the first of a series of lives that can be constructed by an iterative application of GNE to eventually result in L_{nm}:

Transitivity+General Non-Elitism: $L_{nm} \succeq L_1$

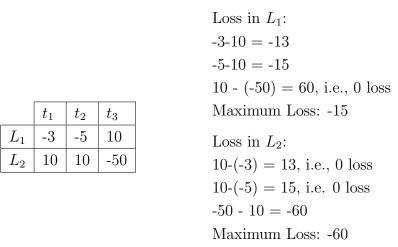|	t_1	t_2	t_3 to t_{n1}	t_{n1+1} to t_{n1+n2}	$t_{n1+n2+1}$ to $t_{n1+n2+n3}$	$\ldots$	$t_{n1+n2+\cdots+nm}$
L_1	100	100	100	1	1	1	1
L_2	99	100	100	99	1	1	1
L_3	99	99	100	99	99	1	1
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
L_{nm}	3	3	3	3	3	3	3

By iteratively decreasing the well-being of the person-stages and making up for it with the well-being of a sufficient number of the remaining person-stages, a series of lives can be constructed such that by Transitivity, $L_{nm} \succeq L_1$. That is, all the person-stages with a well-being level of 100 are first reduced to a well-being level of 99 by making up for it with a sufficient number of remaining person-stages whose well-being is increased from 1 to 99. Then, all of the resulting person-stages with a well-being-level of 99 are reduced to a well-being level of 98 by making up for it with a sufficient number of remaining person-stages whose well-being is increased from 1 to 98. This procedure can be repeated until it results in L_{nm}. However, if the well-being levels are sufficiently fine-grained, it may not seem very plausible that L_{nm} should be at least as good as L_1, even if each judgement in each individual step is. Such continuum cases are of course well known, and are often cited as examples of legitimate intransitive preferences (e.g. Temkin 1996). The present section investigates a theory which escapes the theorem by exhibiting this exact kind of intransitivity, i.e., by endorsing the individual steps generated by General Non-Elitism but inversing the judgement between lives that are sufficiently far away from each other in the series of lives as constructed above.

Denying Transitivity of the better-than relation enables us to block the continuum skeched above: While we can uphold each individual judgement generated by GNE, we can deny that therefore $L_{nm} \succeq L_1$. For instance, we may hold that $L_1 > L_{nm}$, or even earlier in the sequence we can hold that for some x such that $nm > x > 2$ that $L_x > L_{nm}$. However, this intransitivity is of course in need of explanation. When and why does this judgement flip?

The present section examines what I believe to be a very promising proposal to answer this question, namely Minimax Regret. A version of Minimax Regret was introduced by Bruckner (2003), which I will develop further and examine regarding its implications for the questions investigated in this paper. To understand Minimax Regret, it is useful to first introduce Minimax Loss. According to Minimax Loss, we should minimise the

maximal relative loss the different person-stages incur when compared to their temporal counterparts. Consider the following example to illustrate this proposal.

Going to the Dentist With Initial Anxiety

Loss in L_1:

-3-10 = -13

-5-10 = -15

10 - (-50) = 60, i.e., 0 loss

Maximum Loss: -15

	t_1	t_2	t_3
L_1	-3	-5	10
L_2	10	10	-50

Loss in L_2:

10-(-3) = 13, i.e., 0 loss

10-(-5) = 15, i.e. 0 loss

-50 - 10 = -60

Maximum Loss: -60

In the above example, you are again considering whether or not to go to the dentist. Since you are somewhat afraid of dental procedures, you know that if you decide to go, you will feel some anxiety in advance. In L_1, you choose to go, and hence you are somewhat anxious at t_1 (-3), where if you did not go (L_2) you would not be feeling this anxiety at t_1 (10). At t_2, then, you experience the somewhat painful dental procedure in L_1, which you do not in L_2. However, at t_3, an untreated dental health issue will cause you significant pain in L_2 (-50). whereas you are happy and healthy in L_1 (10).

In order to evaluate this choice according to the Minimax Loss principle, we first calculate the loss of each person-stage loss for each life. For L_1, there is a loss of -13 at t_1, a loss of -15 at t_2 and a loss of 0 at t_3. For L_2, there is a loss of 0 at t_1, a loss of 0 at t_2 and a loss of 60 at t_3. Note that positive differences are not counted as "negative loss" (i.e., gain). Minimax Loss now tells us to compare the maximal loss in each life, and choose the life with the lowest maximal loss. This principle thus suggests that we ought to go to the dentist, which converges with our intuitive judgement.

Bruckner's principle is based on similar considerations. However, in order to avoid some obvious counterexamples, he develops Minimax Loss to include forward and backward looking regret, and terms his principle Minimax Regret. He argues that at each point in time, it is not only the loss specific to the person-stage at the time we should consider, but also the (backwards- or forward-looking) regret each person-stage has. Consider the following case as an illustration of this idea.

Regret

	t_1	t_2	t_3
L_1	1	1	30
L_2	20	20	10

The maximal loss (20) in the above case is incurred in L_2 at t_3, such that Minimax Loss would recommend choosing L_2, which is somewhat counterintuitive. Instead, Bruckner defines regret as follows: Each individual person-stage cares not only about their own loss but about the loss of the others as well. Thus, if their past or future person-stage has incurred a loss, this adds regret to the current person-stage. Bruckner introduces a discount rate to calculate regret. That is, each person-stage considers the losses of the other person-stages, but they are discounted based on how temporally removed they are from the specific person-stage. For instance, assume that the the loss of a person-stage is discounted by $\frac{1}{2}$ for each additional temporal index the respective person-stage is removed from the one we are considering. In the above case, then, this would lead to the following regret calculations.

Regret

	t_1	t_2	t_3
L_1	$19 + \frac{19}{2} + \frac{0}{4}$ $= 28.5$	$19 + \frac{19}{2} + \frac{0}{2}$ $= 28.5$	$0 + \frac{19}{2} + \frac{19}{4}$ $= 14.25$
L_2	$0 + \frac{0}{2} + \frac{20}{4}$ $= 5$	$0 + \frac{0}{2} + \frac{20}{2}$ $= 10$	$20 + \frac{0}{2} + \frac{0}{4}$ $= 20$

As illustrated in the matrix above, each person-stage now considers their own loss as well as the losses of the other person-stages in the same life, discounted by how far removed they are. For instance, the person-stage in L_2 at t_1 has its own loss of 0, and adds the loss of the person-stage in L_2 at t_2 discounted by $\frac{1}{2}$, which is also 0, and adds the loss of the person-stage in L_2 at t_3 discounted by $\frac{1}{4}$ (given that it is two temporal indices removed), which is 5. This results in an overall regret of 5. Now, if we minimise the maximal regret understood in this way, we now choose L_2, for the the maximal regret can now be found in L_1.

Bruckner does not consider cases of varying lengths of life. This is a significant problem, given that many complications only arise when lives are unequally long. I will thus build on his theory to extend it to such cases. There are at least two initially plausible ways his theory could be extended. Either we do not count non-existence as a

loss at all, or we treat non-existence equivalent to a well-being level of 0. Not treating non-existence as a loss might be inspired, for instance, by Meacham's harm minimisation view in population ethics (Meacham 2012), an elaborate attempt at spelling out a plausible person-affecting view.

As mentioned in section 3, person-affecting views in population ethics are based on the intuition that while we have a moral reason to make existing people happy, the mere fact that a person's life will be happy does not give us any moral reason to create them. However, this person-affecting feature, i.e., whether our choice will actually affect an existing person, cannot be directly translated to prudence, since prudential issues always affect some person.

For this reason, it is much more plausible to include non-existence as a loss in developing a theory of prudence. To see this more concretely, consider a choice between a very long life with many person-stages with a well-being level of 100 and a very short life with only one person-stage with a well-being level of 100. If non-existence is not counted as a loss, these two lives would be equally good. But this is very implausible. In order to avoid this problem about prudential loss, let us count non-existence as a well-being level of 0.

Extended in this way, Minimax Regret amounts to the following proposal:

> L_x is at least as good as L_y iff $\max_{L_y}\{r_1, r_2, ...r_n\} \geq \max_{L_x}\{r_1, r_2, ...r_m\}$, where r_i denotes loss plus regret as outlined above for the respective person-stage i.

. This way Minimax Regret is able to meet all our desiderata except for Transitivity. It halts the problematic series of lives constructed above based on GNE because the loss each person-stage experiences is small in each individual step but at some point becomes sufficiently large between lives that are far away from each other in the series, such that the judgement flips.

First, Minimx Regret denies Transitivity, which is a substantive conceptual and decision-theoretic cost (cf. Nebel 2018 for a strong argument against intransitivity in spectrum cases). Transitivity is often understood to be an integral property of any betterness relation as well as a plausible axiom in decision theory. Both of these are often justified by coherence considerations or pragmatic arguments, i.e., intransitive betterness judgements or choice behaviour is deemed incoherent or practically self-defeating in some sense. For instance, a prominent type of argument has been from money pumps, i.e., from the fact that intransitive preferences may lead the respective agent to lose an arbitrarily large amount of money. I cannot here provide a comprehensive discussion

24

of the plausibility of denying transitivity, of course. It is worth noting, though, that there have been several recent advances in the exploration of value and decision theory without transitivity (cf., e.g., McClennen 1990, Ahmed 2017, or Thoma 2020). As mentioned above, further arguments in favour of denying transitivity have been mounted based various versions of continuum cases. Intransitive betterness judgements based on multiple values have recently been explored by, e.g., Muñoz (2023). Given this development, the escape route of denying Transitivity is at least worth taking seriously.

Interestingly, we can attempt to plausibilise the intransitive judgements generated by Minimax Regret in two different possible ways: We can either follow Temkin and hold that this reveals the comparative, context-sensitive nature of betterness (1996), or we can follow Muñoz and hold that this reveals that prudence is an instance where betterness is based on a comparison across multiple dimensions (2023). That is, roughly, we can either argue that the notion of "loss" central to Minimax Regret is itself an inherently comparative concept, or, roughly, argue that each individual person-stage-comparison provides a distinct dimension from which to evaluate the lives. While I will not take a stance on this issue here, working out the implications of a theory of prudence on the notion of betterness is an interesting project for further research.

Apart from the bullet of denying transitivity, Minimax Regret also runs into a type of possible applied counterexample, namely that in some cases, it deems a very happy shorter life to be better than an arbitrarily long decently happy life.

Consider the following case. Assume again that our discount rate is $\frac{1}{2}$. Such a Minimax Regret view would recommend a shorter life in the following case.

Short vs Long

	t_1	t_2	$\ldots$	t_n
L_1	452	–	–	–
L_2	1	150	$\ldots$	150

To see why that is, consider the extreme case where n = ∞, i.e., L_2 is actually infinite. Now consider first the overall regret of the person-stage in L_2 at t_2. The loss of the person-stage in L_2 at t_1 is 0, so we only need to consider the forward looking regret. This regret can be calculated as the sum of a sequence.

Sequence: $150, \frac{150}{2}, \frac{150}{4}, \frac{150}{8} \ldots$

This sequence will approach a limit, such that the overall sum of it can be

determined as follows:

$$\frac{150}{1-\frac{1}{2}} = 300$$

The regret of the person-stage in L_2 at t_2, then, is smaller than the regret of the person-stage in L_1 at t_1. Now it is important to note that the regret of the person-stage in L_2 at t_2 is of course not the greatest regret of any of the person-stages in L_2. In fact, the regret increases consistently for every person-stage in L_2. This is because more losses are weighed with a higher discount rate, given that we are looking to both sides (backwards and forwards). For instance, at t_3, the regret sequence for the person-stage includes the term $\frac{150}{2}$ twice, once for the person-stage at t_2 and once for the person-stage at t_4. However, even if *each* of the terms in the above sequence were repeated (which will not even be the case for any particular self, given that the sequence is not infinite in both directions, but we *can* make the number of repetitions of the terms arbitrarily large), such a sequence would still approach a limit of 450. Thus, even for such a self, the regret would not be higher than for the person-stage in L_1.

This, in turn, means that Minimax Regret recommends a shorter life over an infinitely long and decently happy life, which seems rather counterintuitive. This type of result does not exclusively obtain with cases involving infinity, of course. The problem with Bruckner's discount rate is that maximal regret will always ultimately approach a limit, such that a slightly higher level of well-being of one person-stage in a different life is recommended instead of a very long, also very happy life. Depending on the exact parameters of a Minimax Regret principle, i.e., on the exact discount rate and how it is applied to a sequence of person-stages, the difference in well-being between the very long life (such as the above L_2) and the much shorter life with higher well-being (such as the above L_1) is larger or smaller. Yet for any parameter, there will be cases of the type Short vs Long. This is due to the fact that the discount rate introduces certain lexical thresholds, i.e., a threshold above which no further person-stage can raise the regret for other person-stages. This is of course a feature, rather than a bug, when considered in the context of the very repugnant conclusion: We can add as many person-stages with a well-being of 1 as we want without the respective regret for each of these person-stages outweighing the very high levels of well-being in the alternative life. However, if the original life we are considering is actually quite happy, and infinitely long, this feature might turn into a bug.

One possible way of defending Minimax Regret is by setting the parameters of the views such that such Short vs Long cases counterexamples are not too counterintuitive. If we imagine that the difference in well-being between the short and the long life is

sufficiently high, such that the shorter life has to be extraordinarily happy, the examples may not be as troubling as it seems.

Additionally, recall the reinterpretation of the repugnant conclusion in section 4, where we eliminated the possible confounder of one life being much longer than another. I argued that the repugnant conclusion only seems less repugnant in the prudential context if we imagine one life to be very long and the other life quite short. This, I argued, is due to the fact that we might consider length of life itself to be valuable, such that we misinterpret the cases intuitively if we do not adjust the time spans of the cases. Now, if we apply the same reasoning to the above case, it may no longer look so troubling. For instance, if we assume that each period of time lasts for one million years, choosing L_1 is less counterintuitive.

Furthermore, as we shall see in the following section, plausible ways of denying GNE instead runs into very similar counterexamples, such that Short vs Long at least does not favour giving up GNE over giving up Transitivity. Let us thus turn to the escape route of giving up GNE.

4.3   Giving up GNE

Giving up GNE to avoid the spectrum illustrated in the previous section would amount to the following broad picture: Somewhere on the spectrum of lives depicted above, GNE fails, such that a slight decrease in the well-being of one person-stage cannot be successfully outweighed by a respective increase of well-being for any number of person-stages. Call the level of well-being just above this switch x and the level of well-being just below this switch $x-1$. The comparison between the following two lives would thus violate GNE (where where $x - 3 > 1$ and '$\approx$' stands for incommensurability):

Violation of GNE: $L_1 > L_2$ or $L_1 \approx L_2$

	t_1	t_2	t_3	...	t_n
L_1	x	1	1	...	1
L_2	x-1	x-1	x-1	...	x-1

Depending on whether we hold that $L_1 > L_2$ or $L_1 \approx L_2$, this strategy also requires us to give up Completeness. Apart from this, either of these two judgements violates GNE and is sufficient to render the other desiderata consistent. Again, for this to constitute a plausible escape route, however, we need an underlying justification for such a radical break on the spectrum. Why is it that a marginal decrease in well-being for just one

person-stage makes such a big difference in the evaluation of the lives?

The most promising approach to explaining this break, I believe, lies in postulating lexicalities in the nature of well-being itself. The idea that there are higher and lower types of well-being that are not only different in degree, but truly different in kind, goes at least as far back as to Mill (1861, ch. 2). A recent proposal by Nebel, intended as a view on population ethics, which he proposes in "Totalism without Repugnance" (2022), develops this idea. Given the structural similarity between prudence and population ethics, it may thus be very promising to translate his view into the context of prudence. Due to space constraints, I will directly develop Nebel's view in the context of prudence, rather than summarise it in the context of population ethics first.

Nebel introduces a weak lexical priority of high levels of well-being, which he terms important well-being, over low levels of well-being, which he terms trivial well-being, introduces certain thresholds limiting their mutual tradeoffs. Applying this theory to prudence yields the following proposal: There is a certain amount of important well-being Δ, such that if the *difference* in important well-being between too lives exceeds Δ, the life which contains more important well-being is always overall better than the other, no matter how much trivial well-being is present in either of the lives. This, of course, leaves the question of how to evaluate lives where the difference in important well-being does not exceed Δ. Nebel proposes, first, that if one life has both more important and more trivial well-being than another, the former is better than the latter. Second, when important and trivial well-being do not favour the same lives, but the difference in important well-being does not exceed Δ, Nebel introduces a condition to evaluate tradeoffs between important and trivial well-being which compares the ratio of the differences in each category. That is, he introduces a further threshold δ regarding trivial well-being that is used to compare the well-being differences. If none of these conditions apply, the two lives are incommensurable. Overall, a prudential analogue of his view amounts to the following, where i_{L_x} stands for for the sum of important well-being in L_x, and t_{L_x} stands for the sum of trivial well-being in L_x:

> For two sequences of person-stages L_1 and L_2, L_1 is at least as good as L_2 iff,
> 1. $i_{L_1} - i_{L_2} > \Delta$ or
> 2. $i_{L_1} \geq i_{L_2}$ and
>   a. $t_{L_1} \geq t_{L_2}$ or
>   b. $\frac{i_{L_1} - i_{L_2}}{t_{L_2} - t_{L_1}} > \frac{\Delta}{\delta}$
>
> L_1 and L_2 are incommensurable iff none of the above conditions apply.

In population ethics, Nebel calls this the *Lexical Threshold View* (Nebel 2022., 18).[11] and I shall adopt this name for the analogous proposal in prudence. Now let us have a look at how this view violates GNE. Assume that any well-being above -50 and below +50 is considered trivial, while anything else is considered important well-being. Now consider the following comparison of lives:

Violation of GNE: $L_1 > L_2$ or $L_1 \approx L_2$

	t_1	t_2	t_3	$\ldots$	t_n
L_1	50	1	1	$\ldots$	1
L_2	49	49	49	$\ldots$	49

L_1 contains 50 units of important well-being and $n-1$ trivial well-being. L_2 contains 0 important well-being and $49n$ trivial well-being. We can thus note, first, that if $\Delta \leq 50$, then $L_1 > L_2$, which would violate GNE. Second, the amount of important well-being favours L_1 while the amount of trivial well-being favours L_2, such that condition 2.a. does not apply. Lastly, depending on how exactly we set δ, we again get $L_1 > L_2$, which would violate GNE, or none of the conditions apply, such that we get $L_1 \approx L_2$, which likewise violates GNE. Since, it is of course somewhat counterintuitive that L_1 would be better than L_2, setting the parameters such that $L_1 \approx L_2$ makes the violation less problematic.

Supplemented by this theory and underlying interpretation, giving up GNE in the way Nebel proposes looks promising. However, his proposal is faced with three problems: Firstly, it denies Completeness, which is a cost Nebel himself judges to be quite high: Powerful arguments from comparability have recently been mounted in favour of completeness (cf. Dorr, Nebel, and Zuehl 2021; Dorr, Nebel, and Zuehl 2023). Yet completeness has generally been considered easier to deny than transitivity. It may be possible, for instance, to uphold a satisfactory notion of expected utility without the completeness axiom (cf., e.g., McCarthy, Mikkola, and Thomas 2021).

Secondly, it denies the popular assumption that well-being does not exhibit such lexical breaks, which, at least prima facie, seems to be supported by common sense intuition. Lastly, as indicated earlier, it likewise implies that sometimes, an arbitrarily (even infinitely long), decently happy life can never be better than a possibly much shorter very happy life. Let us illustrate this problem based on the Lexical Threshold View. Assume that Δ is the difference in important well-being which makes it such that the life which

---

11. Note that Nebel understands his proposal as a way of determining overall well-being in a population, rather than evaluatively comparing them as I do it here.

contains at least an amount of Δ more important well-being than the other is always judged to be better. Further, assume that any well-being above -50 and below +50 is considered trivial, while anything below/above is considered important well-being. Now consider the following comparison of lives:

Short vs Long: $L_1 > L_2$

	t_1	t_2	$\ldots$	t_n
L_1	$50 + \Delta$	–	–	–
L_2	50	49	$\ldots$	49

In such a comparison, the Lexical Threshold View always judges L_1 to be better, regardless of how large n is. Indeed, even if we set $n = \infty$, the view still suggests that L_1 is better. The same responses given above to defend Minimax Regret apply here too. First, just how counterintuitive these cases are depends on on the exact parameters regarding Δ and important and trivial well-being. That is, we might be able to spell out the parameters in a way that makes the example less counterintuitive. Second, reinterpreting the cases such that both lives are are very long, thereby eliminating confounding intuitions regarding length of life, may likewise contribute to make Short vs Long cases more palatable.

At this point, we have reached an interesting result; namely that giving up Transitivity and giving GNE amount to very similar proposals when supplemented by plausible interpretations. Before comparing their relative advantages, it is worth exploring whether their shared problem may be overcome by giving up GNEP instead.

4.4   Giving up GNEP

If we do not want to accept Short vs Long as outlined in the previous two sections and instead hold on to the judgement that the long life should be better, there might be a way to respond to this challenge that will end up maintaining GNE but violate GNEP. Consider the following negative analogue of Short vs Long:

Short vs Long: $L_2 > L_1$

	t_1	t_2	$\ldots$	t_n
L_1	$-(50 + \Delta)$	–	–	–
L_2	-50	-49	$\ldots$	-49

If we assume that the lexicalities work the same way in the positive and negative domain, the Lexical Threshold View now suggests that the longer life L_2 is better. While this may of course still be counterintuitive, insofar as the person is still suffering in L_2 all the way to t_n, it may still be *less* counterintuitive than the positive version of Short vs Long. That is, we may hold that choosing a very long life of trivial suffering over a short life of important suffering is way less counterintuitive than choosing a short life of important positive well-being over a very long life of positive trivial well-being: At least the former does not involve foregoing an arbitrarily long life.

This view may be supported by both substantively descriptive as well as normative considerations. Klocksiem argues that there is a fundamental difference between what he calls mere discomforts and genuine suffering (2016). That this kind of phenomenology could be real may be more intuitive in the negative domain than in the positive domain: the difference between a mere discomforts, i.e. things we find negative but *bearable*, and genuine suffering, i.e. things we consider genuinely *unbearable* has some intuitive appeal. In contrast, one may argue that there is no such intuitive difference between things we find mildly pleasurable and things we find very pleasurable. This sort of reasoning may be additionally supported by the fact that various moral theories treat minimising very negative suffering as more important than maximising very positive happiness (for reasons beyond the procreation asymmetry). This datum could be explained by a theory which posits a kind of asymmetry between negative and positive well-being itself. I am not aware of a theory that explicitly pursues a theory of prudence or population ethics based on this proposal and it is thus worth sketching here.

Consider the following kind of view: The only point at which well-being exhibits a lexical threshold is between unbearable suffering and everything else. Apart from this threshold, well-being can be traded off against each other in the usual spirit of totalism. Such a view may be even more plausible if it, like the Lexical Threshold View, only exhibits weak lexicality. That is, we may postulate that not just *any* amount of unbearable suffering is prioritised over any amount of other types of well-being, but that there is certain amount Δ such that a difference in unbearable suffering which exceeds Δ always make the life with the higher amount of unbearable suffering worse. Indeed, we can adapt all of Nebel's conditions to construct the *Negative Lexicality View*, where S_{L_x} stands for the sum of unbearable suffering in L_x and w_{L_x} stands for the sum of well-being that is not unbearable suffering in L_x:

For two sequences of person-stages L_1 and L_2, L_1 is at least as good as L_2 iff,

1. $S_{L_1} - S_{L_2} > \Delta$ or
2. $S_{L_1} \geq S_{L_2}$ and
   a. $w_{L_1} \geq w_{L_2}$ or
   b. $\frac{S_{L_1} - S_{L_2}}{w_{L_2} - w_{L_1}} > \frac{\Delta}{\delta}$

L_1 and L_2 are incommensurable iff none of the above conditions apply.

Note that the Negative Lexicality View neither violates GNE nor Non-Repugnance, as defined in above. It does not violate Non-Repugnance since we only took avoiding the very repugnant conclusion as a desideratum, but not the repugnant conclusion. The lexicality in the negative domain of well-being avoids the very repugnant conclusion, even if the absence of such a threshold in the positive domain does not avoid the ordinary repugnant conclusion. And, of course, the view has the additional advantage that it does not recommend a short life over an arbitrarily long life, as giving up transitivity or GNE did.

However, as expected, the Negative Lexicality View violates GNEP. Consider the following case. Assume again that well-being of -50 or lower constitutes unbearable suffering.

Violating GNEP: $L_1 > L_2$ or $L_1 \approx L_2$

	t_1	t_2	$\dots$	t_n
L_1	−49	1	$\dots$	1
L_2	-50	1000	$\dots$	1000

The sum of unbearable suffering in L_1 is 0, and the sum of other well-being in L_1 is $n-50$. The sum of unbearable suffering in L_2 is -50, and the sum of other well-being in L_2 is $1000n - 1000$. First, it holds again that if $\Delta \leq 50$, then $L_1 > L_2$, which violates GNEP. Second, the sum of unbearable suffering favours L_1 while the sum of other well-being favours L_2, such that condition 2.a. does not apply. Lastly, depending on how exactly we set δ, we again get $L_1 > L_2$, which would violate GNEP, or none of the conditions apply, such that we get $L_1 \approx L_2$, which likewise violates GNEP. In contrast to the analogous case we examined when giving up GNE, it is even more implausible to set the parameters such that $L_1 > L_2$. After all, the difference in other well-being is vastly larger than the difference in unbearable suffering. This is fortunate since it is intuitively more

plausible that $L_1 \approx L_2$ than that $L_1 > L_2$. It is thus plausible that the Negative Lexicality View only ever exhibits incommensurability when violating GNEP. In particular, since GNEP only ever involves big well-being differences, while GNE can also involve small well-being differences (since for GNE the added well-being must just be higher by at least one unit), it is thus *more plausible* for the Negative Lexicality View to only ever violate GNEP based on incommensurability, than it is for the Lexical Threshold View to only ever violate GNE based on incommensurability. This may constitute an additional advantage of the Negative Lexicality View over the Lexical Threshold View. And insofar as avoiding the negative version of Short vs Long is indeed less important than avoiding the positive version, giving up GNEP constitutes a serious alternative escape route.

5   Conclusions and Implications

Having examined four possible escape routes, the question of which one is most plausible remains. While I do not take a definitive stance on this question, it is useful to highlight the central costs to each of them again.

	Non-Repugnance	**Transitivity**	**GNE**	**GNEP**
Theory	Totalism	Minimax Regret	Lexical Threshold View	Negative Lexicality View
Structural Problems	-	Intransitivity	Incompleteness	Incompleteness
Substantive Bullets	Very Repugnant Conclusion	Short vs Long	Short vs Long, GNE Violation	Negative Short vs Long, GNEP Violation
Substantive Assumptions	Full Intrapersonal Compensation	Intrapersonal Regret	Lexical Well-Being	Lexical Well-Being

There are several dividing lines between the four routes. In terms of substantive bullets, it seems to me, the Negative Lexicality View is able to balance best both the intuition that long - even somewhat dull - lives do have high value and the intuition that not every terrible torturous stage in one's life can be compensated by this value. In terms of substantive assumptions, however, positing that, as a matter of prudence, person-stages have some regard for the other person-stages in the form of regret seems to be a relatively weak assumption compared to the idea that well-being, by its very nature, is lexical. This may in turn favour Minimax Regret. However, giving up Transitivity, as Minimax Regret does, may be too high a decision-theoretic cost. Note, however, that the Lexical Threshold View and the Negative Lexicality View likewise introduce

intransitive *choice-behaviour*. If incommensurability is translated into indifferent choice-behaviour, i.e., choosing either way is permissible, then the Lexical Threshold View and the Negative Lexicality View will recommend intransitive choice-behaviour, even if the underlying axiology is transitive. Insofar as we want to avoid any intransitivity, even choice-behaviourally, developing the Lexical Threshold View or the Negative Lexicality View in way that avoids this may be desirable.

Interestingly, the choice in theory of prudence may have ramifications for population ethics and moral theory more broadly. First, if we accept either the Lexical Threshold View or the Negative Lexicality View, this plausibly grounds analogous solutions for population ethics: If well-being by its very nature exhibits these lexicalities, it is plausible that this structure transfers to its aggregates. Indeed, the lexical nature of well-being may ground partial aggregation more broadly. For instance, Scanlons famous Transmitter Room Case, whereby one person experiences terrible electricity shocks for the trivial benefit of billions can be explained and grounded by lexical well-being (Scanlon 1998, p. 235). Such a unified solution may be theoretically attractive.

If we adopt Minimax Regret instead, this has ramifications for decision theory and the nature of betterness more generally. It would lend support to decision theory without transitivity as an axiom and intransitive conceptions of betterness. This, in turn, would suggest that it may be similarly interesting to further investigate whether there may be an analogous view for population ethics. After all, if prudential betterness is intransitive, why should population betterness not be? Such a view could be based on Meacham's moral counterparts, supplemented by some imposed ordering on the population. For instance, we may order populations and their counterparts in accordance with how badly off the individuals are, thereby introducing an additional component of moral priority.

The difficulty in finding an adequate theory of prudence as I have sketched it here might ultimately also point to an inadequate understanding of well-being underlying my investigations. I characterised prudence as a matter of separable and comparable levels of well-being that can be intrapersonally aggregated. However, it may turn out that this understanding is misguided. For instance, it may ultimately be the case that well-being itself can only be assessed atemporally from the perspective of the overall life, rather than at any given moment or in a given time period. Further investigating the plausibility of this understanding of prudence might thus be a further important avenue of research.

Appendix

I shall show, firstly, how the formal framework of population ethics can be adapted to prudence. Then, I shall state the constraints from section 4 more formally within this framework. Finally, we can rehearse the proof for the impossibility theorem in an analogous manner.

In doing so, I shall partly rely on Thomas' helpful formal framework (2016). He introduces a finite set of well-being levels:

$$\mathbb{W} : \{Z, Z+1, \ldots, -2, -1, 0, 1, 2, \ldots, A, A+1, A+2\},$$

where Z corresponds to an abominably low well-being level, 0 to a neutral level, and A and above to extremely high well-being levels. Furthermore, he makes use of a well-being distribution which maps the set of well-being levels onto the non-negative integers. This integer represents how many people exist in the population that have the respective level of well-being (Thomas 2016, 2).

For our purposes, we can adapt this framework by taking on the set of well-being levels, and define the well-being distribution such that it maps the level of well-being to the number of *person-stages* which have the respective level of well-being. Note, however, that in this way, we lose the temporal ordering of the sequences of person-stages. This is not a problem for our purposes, given that the proofs we aim to replicate do not make reference to the ordering of the sequence. However, if and to the extent that we would eventually want to include the ordering in finding the right principle of prudence, a respective formalisation will be needed. This is especially important considering the fact that some desiderata we have introduced are especially plausible for prudence given a certain temporal ordering. I will leave this up for further research, however, and ignore the ordering for now.

Thomas then uses the following notation to denote populations: n[x] denotes that n number of people are at well-being level x. We can construct populations with varying levels of welfare from this by adding terms of this sort. For instance, n[x] + m[y] denotes the population of n people at well-being level x and m people at well-being level y (Thomas 2016, 2). Given our definition of the well-being distribution, this notation is semantically adapted to denote the number of person-stages at the particular well-being level. n[x] + m[y] thus denotes a sequence of person-stages where n number of person-stages are at well-being level x and m number of person-stages are at well-being level y, where these person-stages can be ordered in any possible way.

35

Desiderata for Prudence

Completeness: For any two lives L_x and L_y, $L_x \succeq L_y$ or $L_y \succeq L_x$ (or both).

Transitivity: For any three lives L_x, L_y and L_z, if $L_x \succeq L_y$, and $L_y \succeq L_z$, then $L_x \succeq L_z$.

Egalitarian Dominance: For any well-being levels $x < y$ and any number of person-stages $n \in \mathbb{N}$, $\quad n[y] \succ n[x]$

Dominance Addition: For any well-being levels $x > y$ and $z > 0$, and any number of person-stages $m, n \in \mathbb{N}$, $\quad m[x] + n[z] \succeq m[y]$

General Non-Elitism: For all $x, z \in \mathbb{W}$, with $x > z + 1$, there exists $G(x, z) \in N$ such that, for any life L,

$$S + 1[x - 1] + G(x, z)[x - 1] \succeq S + 1[x] + G(x, z)[z]$$

General Non-Extreme Priority: For any $z \in \mathbb{W}$, there exists $G(z) \in \mathbb{N}$ such that, for any well-being level $x \geq A$, any welfare level $0 < y \leq 3$, and any life L,

$$S + 1[z - 1] + G(z)[x] \succeq S + 1[z] + G(z)[y].$$

Non-Repugnance: It is not the case that for any $m, n \in N$, and any well-being level $z < 0$, there exists $G(m, n, z) \in N$ such that
$$m[z] + G(m, n, z)[3] \succ n[A].$$

Impossibility Theorem

For given $m, n \in N$, given $z < 0$, and some B, C:

(1) $n[A + 1] > n[A]$

Egalitarian Dominance: $L_2 > L_1$

	t_1	t_2	...	t_n
L_1	A	A	...	A
L_2	A+1	A+1	...	A+1

(2) $n[A + 2] + m[1] + B[1] + C[1] \geq n[A + 1]$

Dominance Addition: $L_3 \geq L_2$

	t_1	...	t_n	t_{n+1}	...	t_{n+m}	t_{n+m+1}	...	t_{n+m+B}	$t_{n+m+B+1}$	...	$t_{n+m+B+C}$
L_2	A+1	...	A+1	-	-	-	-	-	-	-	-	-
L_3	A+2	...	A+2	1	...	1	1	...	1	1	...	1

(3) $n[A + 2] + m[z] + B[A + 2] + C[1] \geq n[A + 2] + m[1] + B[1] + C[1]$

General Non-Extreme Priority: $L_4 \geq L_3$

	t_1	...	t_n	t_{n+1}	...	t_{n+m}	t_{n+m+1}	...	t_{n+m+B}	$t_{n+m+B+1}$	...	$t_{n+m+B+C}$
L_3	A+2	...	A+2	1	...	1	1	...	1	1	...	1
L_4	A+2	...	A+2	z	...	z	A+2	...	A+2	1	...	1

Note that in order to generate the overall judgement that $L_4 \geq L_3$, we must go through several application of General Non-Extreme Priority iteratively. That is, we apply it step by step to individual person-stages and individual decreases of their well-being, making up for them by a sufficient number of person-stages with a well-being level of A+2. The number B is determined by however many person-stages are necessary to iteratively make up for m person-stages with well-being level z. Ultimately, we arrive at a life of the sort of L_4. By transitivity, $L_4 \geq L_3$ holds based on this procedure.

The same iterative procedure is necessary in the following step, where C is determined by n and B.

(4) $n[3] + m[z] + B[3] + C[3] \geq n[A + 2] + m[z] + B[A + 2] + C[1]$

General Non-Elitism: $L_5 \geq L_4$

	t_1	...	t_n	t_{n+1}	...	t_{n+m}	t_{n+m+1}	...	t_{n+m+B}	$t_{n+m+B+1}$	...	$t_{n+m+B+C}$
L_4	A+2	...	A+2	z	...	z	A+2	...	A+2	1	...	1
L_5	3	...	3	z	...	z	3	...	3	3	...	3

Thus, given transitivity, $G(m, n, z) = n + B + C$ is the number which generates the very repugnant conclusion, since we have shown that we can always find a number according to which $L_5 \succ L_1$.

References

Ahmed, Arif. 2017. "Exploiting Cyclic Preference." *Mind* 126 (504): 975–1022.

Arrhenius, Gustaf. 2000. "An Impossibility Theorem for Welfarist Axiologies." *Economics and Philosophy* 16 (2): 247–266.

———. 2003. "The Very Repugnant Conclusion." In *Logic, Law, Morality: Thirteen Essays in Practical Philosophy in Honour of Lennart Åqvist,* edited by Krister Segerberg and Ryszard Sliwinski, 29–44. Uppsala.

———. 2009. "One More Axiological Impossibility Theorem." In *Logic, Ethics and All That Jazz. Essays in Honour of Jordan Howard Sobel,* edited by Lars-Göran Johansson, Jan Österberg, and Ryszard Sliwinski, 23–37. Uppsala: Uppsala Philosophical Studies.

———. 2011. "The Impossibility of a Satisfactory Population Ethics." In *Descriptive and Normative Approaches to Human Behavior, Advanced Series on Mathematical Psychology,* edited by Hans Colonius and Ehtibar N. Dzhafarov.

Bentham, Jeremy. 1780. *An Introduction to the Principles of Morals and Legislation.* Edited by J. H. Burns and H. L. A. Hart. New York: Dover Publications.

Bricker, Phillip. 1980. "Prudence." *Journal of Philosophy* 77 (7): 381–401.

Brink, David O. 2003. "Prudence and Authenticity: Intrapersonal Conflicts of Value." *Philosophical Review* 112 (2): 215–245.

Broome, John. 2017. "Utilitarian Metaphysics?" In *Weighing Goods,* 224–240. Wiley.

Bruckner, Donald W. 2003. "A Contractarian Account of (Part of) Prudence." *American Philosophical Quarterly* 40 (1): 33–46.

Dorr, Cian, Jacob M. Nebel, and Jake Zuehl. 2021. "Consequences of Comparability." *Philosophical Perspectives* 35 (1): 70–98.

———. 2023. "The Case for Comparability." *Noûs* 57 (2): 414–453.

Dorsey, Dale. 2018. "Prudence and Past Selves." *Philosophical Studies* 175 (8): 1901–1925.

———. 2021. *A Theory of Prudence.* Oxford: Oxford University Press.

Huckfeldt, Vaughn. 2011. "Prudence, Commitments and Intertemporal Conflicts." *Theoria* 77 (1): 42–54.

Hurley, Susan L. 1989. *Natural Reasons: Personality and Polity.* New York: Oxford University Press.

Klocksiem, Justin. 2016. "How to Accept the Transitivity of Better Than." *Philosophical Studies* 173 (5): 1309–1334.

McCarthy, David, Kalle Mikkola, and Teruji Thomas. 2021. "Expected utility theory on mixture spaces without the completeness axiom." *Journal of Mathematical Economics* 97:102538.

McClennen, Edward Francis. 1990. *Rationality and Dynamic Choice: Foundational Explorations.* Cambridge, England: Cambridge University Press.

Meacham, Christopher J. G. 2012. "Person-affecting views and saturating counterpart relations." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 158 (2): 257–287. Accessed June 8, 2023.

Mill, John Stuart. 1861. *Utilitarianism.*

Muñoz, Daniel. 2023. "Sources of Transitivity." *Economics and Philosophy* 39 (2): 285–306.

Nebel, Jacob M. 2018. "The Good, the Bad, and the Transitivity of *Better Than.*" *Noûs* 52 (4): 874–899.

———. 2019. "An Intrapersonal Addition Paradox." *Ethics* 129 (2): 309–343.

———. 2022. "Totalism Without Repugnance." In *Ethics and Existence: The Legacy of Derek Parfit,* edited by Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan, 200–231. Oxford: Oxford University Press.

Parfit, Derek. 1984. *Reasons and Persons.* Oxford University Press.

Paul, L. A. 2014. *Transformative Experience.* Oxford University Press.

Pettigrew, Richard. 2019. *Choosing for Changing Selves.* Oxford, UK: Oxford University Press.

Price, Bruce W. 2002. "The Worthwhileness Theory of the Prudentially Rational Life." *Journal of Philosophical Research* 27:619–639.

Rawls, John. 1971. *A Theory of Justice: Original Edition.* Belknap Press.

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

Sidgwick, Henry. 1901. *Methods of Ethics.* Kaplan.

Taurek, John. 1977. "Should the Numbers Count?" *Philosophy and Public Affairs* 6 (4): 293–316.

Temkin, Larry S. 1996. "A Continuum Argument for Intransitivity." *Philosophy and Public Affairs* 25 (3): 175–210.

Thoma, Johanna. 2020. "Preference Cycles and the Requirements of Instrumental Rationality."

Thomas, Teruji. 2016. "Reconstructing Arrhenius's Impossibility Theorems."

Velleman, J. David. 1991. "Well-Being And Time." *Pacific Philosophical Quarterly* 72 (1): 48–77.