Free & Equal | A Journal of Ethics and Public Affairs

Patrick McKee 2025 The Unhappy Conclusion. *Free & Equal: A Journal of Ethics and Public Affairs*, 1(2): 445–471. DOI: https://doi.org/10.16995/fe.18353

Open Library of Humanities

THE UNHAPPY CONCLUSION

Patrick McKee, Philosophy, Massachusetts Institute of Technology, mckee@mit.edu

I argue that it is better to live an extremely long, drab life than a happy life of normal length. I rely on four premises, concerning (1) the separability of well-being in time, (2) the circumstances in which we should prolong someone's life, (3) the shape-of-a-life hypothesis, and (4) tradeoffs between different levels of well-being within a life. I show that the conclusion has implications for the Millian lexical superiority view and for population ethics. It may also have implications for the well-being of animals in captivity and artificially intelligent welfare subjects.

Open Library of Humanities

Free & Equal A Journal of Ethics and Public Affairs

https://doi.org/10.16995/fe.18353

Volume 1, Issue 2 (2025): 445–471

ARTICLE

THE UNHAPPY CONCLUSION

Patrick McKee

I argue that it is better to live an extremely long, drab life than a happy life of normal length. I rely on four premises, concerning (1) the separability of well-being in time, (2) the circumstances in which we should prolong someone's life, (3) the shape-of-a-life hypothesis, and (4) tradeoffs between different levels of well-being within a life. I show that the conclusion has implications for the Millian lexical superiority view and for population ethics. It may also have implications for the well-being of animals in captivity and artificially intelligent welfare subjects.

I. Introduction

I will argue that a long life is better than a happy life. More precisely, I will argue for the,

Unhappy Conclusion For any finite happy life *a*, there is a finite life *z* such that every day in *z* is drab and low in well-being, but *z* offers higher lifetime well-being than *a*.[1]

J. M. E. McTaggart first proposed the *Unhappy Conclusion*.[2] He thought it true, on the grounds that it followed from an additive picture of value over time. But he also thought many would find it "repugnant"; I call it "unhappy" to convey the same sort of reaction. Most recent commentators have thought it false.[3]

1. Though I formulate the *Unhappy Conclusion* in terms of days, it could instead be formulated in terms of hours, minutes, or seconds, without affecting the argument. It could also be formulated in terms of durationless moments, but that would introduce technical complexity related to the infinity of moments in a life which it would be simpler to avoid.

2. J. M. E. McTaggart, *The Nature of Existence*, Vol. 2 (Cambridge University Press, 1927), 452–53.

3. Notably, James Griffin, *Well-Being: Its Meaning, Measurement, and Moral Importance* (Clarendon Press, 1986), chapter 5; Douglas Portmore, "Does the Total Principle Have Any Repugnant Implications?" *Ratio* 12, no. 1 (1999): 80–98, https://doi.org/10.1111/1467-9329.00078; Derek Parfit, "Overpopulation and the Quality of Life," in *The Repugnant Conclusion: Essays on Population Ethics*, eds. Jesper Ryberg and Torbjörn Tännsjö (Kluwer Academic Publishers, 2004), 7–22, https://doi.org/10.1007/978-1-4020-2473-3_2; and Jacob Nebel, "Totalism Without Repugnance," in *Ethics and Existence: The Legacy of Derek Parfit*, eds. Jeff McMahan, Tim Campbell,

If true, the *Unhappy Conclusion* is important for several reasons. First, it is a foundational claim about the structure of well-being over time. It implies that any decrease in quality of life, down to the point where quality is barely positive, can be compensated by a sufficient increase in quantity of life. It may thus inform some practical decisions. The decision to keep an animal in captivity can involve a dramatic quality-quantity tradeoff. A cat who lives indoors, safe from cars and coyotes, can expect to live much longer than one who lives outdoors. But she cannot hunt or stake out a territory, two things that, for a cat, make life worth living. The prospect of artificially intelligent welfare subjects also compels us to evaluate long, modest-quality lives. These beings will have no biological limit on their lifespan, but they might, for the sake of our safety, be precluded from enjoying important goods like autonomy and embodiment.[4]

Second, the *Unhappy Conclusion* bears on the structure of well-being across different kinds of goods. John Stuart Mill popularized the view that there are (at least) two kinds of goods, such that some amount of a higher good is better for us than an arbitrarily large amount of a lower good.[5] Call this the *lexical superiority view about well-being*. The higher goods have been taken to include friendship, meaningfulness, and the appreciation of beauty; the lower, mild physical pleasure. But one cannot enjoy an arbitrarily large amount of mild physical pleasure all at once. So, in effect, the lexical superiority view about well-being says that some amount of a higher good is better for us than an arbitrarily long time of enjoying only lower goods. This is more or less the negation of the *Unhappy Conclusion*.

Third, the *Unhappy Conclusion* has implications for population ethics. It is an intrapersonal analog of the

James Goodrich, and Ketan Ramakrishnan (Oxford University Press, 2022), 200–31, https://doi.org/10.1093/oso/9780192894250.003.0009.

4. Carl Shulman and Nick Bostrom note that the potentially enormous lifespan of an AI means that killing it (if that is the right word) can do it enormous harm. See Carl Shulman and Nick Bostrom, "Sharing the World with Digital Minds," in *Rethinking Moral Status*, eds. Steve Clarke, Hazem Zohny, and Julian Savulescu (Oxford University Press, 2021), 306–26, https://doi.org/10.1093/oso/9780192894076.003.0018. On the prospect of artificially intelligent welfare subjects, see also Robert Long et al., "Taking AI Welfare Seriously," preprint, arXiv, November 4, 2024, https://doi.org/10.48550/arXiv.2411.00986 and Simon Goldstein and Cameron Domenico Kirk-Giannini, "AI Wellbeing," *Asian Journal of Philosophy* 4 (2025): 1–22, https://doi.org/10.1007/s44204-025-00246-2.

5. John Stuart Mill, *Utilitarianism*, 7th ed. (Longmans, 1879), chapter 2. For a general discussion of lexical superiority views and their history, see Gustaf Arrhenius, "Superiority in Value," *Philosophical Studies* 123, nos. 1-2 (2005): 97–114, https://doi.org/10.1007/s11098-004-5223-0.

Repugnant Conclusion For any finite happy population *A*, there is a finite population *Z* such that every person in *Z* lives a life that is drab and low in well-being, but *Z* is better than *A*.[6]

Given the analogy, one might suspect that the two conclusions stand or fall together. But notice that, whereas the *Unhappy Conclusion* concerns the aggregation of one kind of value (well-being), the *Repugnant Conclusion* concerns the way in which two kinds of value (well-being and general good) relate. As we will see, that gives us more latitude to resist the *Repugnant Conclusion*. That said, one popular response to the *Repugnant Conclusion* holds that certain higher goods, like friendship, meaningfulness, and the appreciation of beauty, contribute more to the general good of a population than an arbitrarily large amount of lower goods, like mild physical pleasure. Call this the *lexical superiority view about general good*. I will suggest that this view probably stands or falls with the lexical superiority view about well-being—and so with the *Unhappy Conclusion*.

I will argue for the *Unhappy Conclusion* from four premises. The premises are jointly weaker than McTaggart's idea that value is additive over time. In addition to the four premises, I will assume that well-being comparisons are transitive. The transitivity of *is at least as good as*, and indeed of any predicate of the form *is at least as F as*, seems to me a conceptual truth, though there is a lively debate about it to which I cannot do justice here.[7] However, I will not assume that well-being comparisons are complete. It is consistent with my argument that some pairs of lives be incommensurable, that is, such that neither is at least as good as the other.[8]

6. Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), chapter 17. I have modified Parfit's original formulation to concern well-being rather than quality of life and to specify that the lives in *Z* are drab. Parfit takes the latter specification to make the conclusion especially repugnant: see Parfit, "Overpopulation and the Quality of Life." Others have noted the analogy between the *Unhappy Conclusion* and the *Repugnant Conclusion*: see Griffin, *Well-Being*, chapter 5; Tyler Cowen, "Normative Population Theory," *Social Choice and Welfare* 6, no. 1 (1989): 33–43, https://doi.org/10.1007/BF00433361; Portmore, "Does the Total Principle Have Any Repugnant Implications?," 84–87; John Broome, *Weighing Lives* (Oxford University Press, 2004), https://doi.org/10.1093/019924376X.001.0001; Parfit, "Overpopulation and the Quality of Life"; and Nebel, "Totalism Without Repugnance."

7. For arguments against transitivity, see Larry Temkin, "Intransitivity and the Mere Addition Paradox," *Philosophy and Public Affairs* 16, no. 2 (1987): 138–87 and Stuart Rachels, "Counterexamples to the Transitivity of Better Than," *Australasian Journal of Philosophy* 76, no. 1 (1998): 71–83, https://doi.org/10.1080/00048409812348201. For a critical response, see Jacob Nebel, "The Good, the Bad, and the Transitivity of Better Than," *Noûs* 52, no. 4 (2018): 874–99, https://doi.org/10.1111/nous.12198.

8. Thus I assume that the relation *is at least as good as* ($\geqslant$) is a non-strict partial order on the set of possible lives.

The plan is as follows. In §§II–V, I will introduce and defend the four premises. In §VI, I will prove the *Unhappy Conclusion* from the premises. In §VII, I will discuss the implications for population ethics.

II. Separability

My first premise is

Separability If two lives of equal length are equally good on some days, then their relative ranking by lifetime well-being depends only on how good they are on the other days.[9]

Imagine that two people have equally good childhoods. *Separability* says that to know which of these people has a better life as a whole, we only need to know who has a better adulthood: we don't need to know whether they both had happy childhoods or miserable childhoods. Or imagine a philosopher who looks back on her life and wonders whether she would have been better off as a physicist. It would be odd for her to think that she was better off as a philosopher, but that, had her annual beach vacations been less enjoyable, she would have been better off as a physicist. The relative ranking of her philosopher-life and her physicist-life shouldn't depend on the value of vacation days that are equally good in both lives. This is what *Separability* implies.

Some philosophers deny *Separability* on the ground that there are *timeless goods*: goods that affect well-being in a lifetime but not on any particular day. If there are such goods, then the relative ranking of two lives by lifetime well-being can depend on more than how good they are on each day, and, *a fortiori*, on more than how good they are on the days on which they differ in well-being. David Velleman argues that the quality of one's life story is a timeless good.[10] He imagines someone who is unhappily married and can choose either to fix her marriage or to divorce and remarry. Even if the options are equally good for her on each day, Velleman holds, fixing the marriage will give her the better life as a whole. That's because "a dead-end relationship blots the story of one's life in a way that marital problems don't if

9. For extra clarity, I'll formalize certain claims in footnotes. I'll use Latin letters (x, y, z, w) to denote lives, and subscripted Latin letters to denote time periods within lives. For example, if t denotes the first day of a life, x_t denotes the first day of life x. *Separability* can be formalized as follows: Let x, y, z and w be any four lives of equal length, and let U and V be any two sets of days that partition the lives; if, for all days $u \in U$, $x_u \sim y_u$ and $z_u \sim w_u$, and for all days $v \in V$, $x_v \sim z_v$ and $y_v \sim w_v$, then $x \succcurlyeq y$ and just in case $z \succcurlyeq w$. It is what John Broome calls "strong separability," in the setting of well-being over time: see Broome, *Weighing Lives*.

10. The *locus classicus* of this view is David Velleman, "Well-Being and Time," *Pacific Philosophical Quarterly* 72, no. 1 (1991): 48–77, https://doi.org/10.1111/j.1468-0114.1991.tb00410.x.

they lead to eventual happiness."[11] Others have proposed achievement and meaningfulness as timeless goods.[12]

It is worth noting that we can accept that these are goods without accepting that they are timeless. The value of these goods might devolve onto the days containing the events that constitute them. The idea would be that, if the unhappy days of a troubled marriage will lead to eventual happiness, they are not as bad as they seem at the time. Jeff McMahan defends this devolution view, and others are sympathetic to it.[13] The devolution view is consistent with *Separability*, which permits the value of a given day to depend (non-causally) on what happens on other days. Alternatively, the putative timeless goods might be good instrumentally, since one can enjoy looking back on one's life story.[14] Or they might be good evidentially, since a good life story can provide evidence of desires satisfied.[15] These views are also consistent with *Separability*.

There are more modest non-separable views. One might hold, with the timeless goods view, that certain events or goods make a context-dependent

11. Velleman, "Well-Being and Time," 55.

12. See, respectively, Roger Crisp, "Utilitarianism and the Life of Virtue," *Philosophical Quarterly* 42, no. 167 (1992): 139–60, https://doi.org/10.2307/2220212, 149–50 and Antti Kauppinen, "Meaningfulness and Time," *Philosophy and Phenomenological Research* 84, no. 2 (2012): 345–77, https://doi.org/10.1111/j.1933-1592.2010.00490.x, 374. See also Griffin, *Well-Being*, 34–37 and Elizabeth Anderson, *Value in Ethics and Economics* (Harvard University Press, 1993), 34–35.

13. See Jeff McMahan, *The Ethics of Killing: Problems at the Margins of Life* (Oxford University Press, 2002), https://doi.org/10.1093/0195079981.001.0001, 176–77; Broome, *Weighing Lives*, 45–8; Douglas Portmore, "Welfare, Achievement, and Self-Sacrifice," *Journal of Ethics and Social Philosophy* 2, no. 2 (2007): 1–29, https://doi.org/10.26556/jesp.v2i2.22; Dale Dorsey, "The Significance of a Life's Shape," *Ethics* 125, no. 2 (2015): 303–30, https://doi.org/10.1086/678373, 323–29; and possibly Kauppinen, "Meaningfulness and Time," 375. The devolution view is especially popular among friends of the desire-satisfaction theory of well-being. Suppose you satisfy at one time a desire you have only at a different time. Most accept that, if this is good for you, then it's good for you at some time. But it's controversial which time this is. See Chris Heathwood, "The Problem of Defective Desires," *Australasian Journal of Philosophy* 83, no. 4 (2005): 487–504, https://doi.org/10.1080/00048400500338690; Ben Bradley, *Well-Being and Death* (Oxford University Press, 2009), https://doi.org/10.1093/acprof:oso/9780199557967.001.1, 18–30; Dale Dorsey, "Desire-Satisfaction and Welfare as Temporal," *Ethical Theory and Moral Practice* 16, no. 1 (2013): 151–71, https://doi.org/10.1007/s10677-011-9315-6; Donald Bruckner, "Present Desire Satisfaction and Past Well-Being," *Australasian Journal of Philosophy* 91, no. 1 (2011): 15–29, https://doi.org/10.1080/00048402.2011.632016; and Eden Lin, "Asymmetrism about Desire Satisfactionism and Time," in *Oxford Studies in Normative Ethics, Vol. 7*, ed. Mark Timmons (Oxford University Press, 2017), https://doi.org/10.1093/oso/9780198808930.003.0009; though, for a dissenting view, see Duncan Purves, "Desire Satisfaction, Death, and Time," *Canadian Journal of Philosophy* 47, no. 6 (2017): 799–819, https://doi.org/10.1080/00455091.2017.1321910.

14. Fred Feldman, *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism* (Clarendon Press, 2004), chapter 6, https://doi.org/10.1093/019926516X.003.0007.

15. Thanks to Bernard Reginster for this point.

contribution to the value of a life, but also allow that these events or goods make a context-independent contribution to the value of a day. For example, some good, like pleasure, might be held to make a decreasing marginal contribution to lifetime well-being. On this view, a trip to the ballgame contributes a fixed amount to the day on which it occurs but contributes less to lifetime value the more pleasure the life contains elsewhere. Conversely, some good might be held to make an increasing marginal contribution to lifetime well-being. The *weak* version of the lexical superiority view says that a higher good makes a lexical contribution to lifetime well-being only if one enjoys enough of it. Perhaps a decade of friendship is lexically superior to pleasure, but a day is not. (By contrast, the *strong* version of the lexical superiority view says that even a day of friendship is lexically superior to pleasure; this version of the view is consistent with *Separability*.)[16] Or one might hold that it is important for a life to contain the right balance of pleasure and friendship.[17]

I will now give two arguments against non-separable views. Both arguments exploit the fact that well-being in a day and well-being in a lifetime can be bridged with well-being in intermediate periods. One such period occupies a familiar place in our prudential thinking: *the future*.[18] I will show that non-separable views conflict with each of two appealing principles about the future.

My first argument appeals to:

Future Dominance If x and y are two lives of equal length, x is at least as good as y tomorrow, and x is at least as good as y in the future after tomorrow, then x is at least as good as y in the future as a whole.

I will defend the principle with a slightly modified version of Velleman's case:

16. The "weak" vs. "strong" terminology comes from Arrhenius, "Superiority in Value;" see also Griffin, *Well-Being*, chapter 5 and Nebel, "Totalism Without Repugnance." One might also take the relevant "good" to be a certain level of daily well-being, however achieved, rather than any good in particular.

17. I thank an anonymous referee for encouraging me to address other non-separable views and for suggesting ways in which they might be elaborated.

18. It may be worth noting that Velleman, "Well-Being and Time," draws two distinctions: (a) between well-being in a moment and well-being in an extended period of time, and (b) between well-being in shorter extended periods such as days and well-being in the longer extended periods they comprise (see, for example, ibid., 48). Since the *Unhappy Conclusion* concerns days vs. lifetimes, I'm focused on distinction (b). But, for what it's worth, it doesn't seem to me that distinction (a) reflects a difference in kind: if there is such a thing as momentary well-being, surely it isn't all that different than well-being in a second.

Marriage Tomorrow, Alex will fight with her spouse for the last time. The day after tomorrow, she'll decide whether to fix the marriage or to divorce. If she divorces, she'll be slightly happier in every day of the period starting the day after tomorrow.[19]

On the timeless goods view, fixing the marriage is better for Alex's life as a whole, even though it is not better on any day. Indeed, it is worse on every day starting the day after tomorrow. And, presumably, it is worse in the period starting the day after tomorrow, considered as a whole, because that period contains none of the struggle that fixing the marriage would redeem.

What about Alex's future as a whole? I'm imagining that, on the timeless goods view, Alex's future as a whole will be better if she fixes the marriage. That is because her future *does* contain some of the struggle that fixing the marriage would redeem.[20] Alternatively, the proponent of the timeless goods view could insist that fixing the marriage only improves Alex's *life as a whole*, and no periods within it—not even the period starting when she is one day old. But this seems an implausible place to draw the line (and it does not seem to be Velleman's position).

The proponent of timeless goods should, then, take the case to be a counterexample to *Future Dominance*. She should advise Alex as follows. "Look," she should say, "if you fix the marriage, tomorrow will be just as bad for you, and the future after tomorrow will be worse. Still, your future as a whole will be better." This advice sounds like it cannot be right. It even sounds borderline confused. Alex can rightly respond: "But *when* will I be better off?" This is not just table-pounding, because the answer "In your life as a whole" is no response to it.[21] We often think prudentially about the future, and it seems to me that the advice offered here runs counter to the way in which we think about it.

19. I have modified Velleman's case to make Alex slightly happier after she divorces. This makes the case more vivid. The proponent of timeless goods should still accept that Alex's life as a whole will be better if she fixes the marriage, since timeless goods are presumably meant to be more than tiebreakers between lives of otherwise exactly equal value. (Otherwise, how could we get an intuitive grip on them?)

20. If redeeming one day of struggle doesn't seem enough, imagine the marriage lasts only a few bitter days. Or imagine Alex is to endure a year, rather than a day, of struggle before deciding whether to divorce: then the case will support a version of *Future Dominance* concerning years rather than days, but the structure of the argument will be unchanged. Incidentally, tweaking the case in this way shows that *Future Dominance* would be just as plausible if it concerned years rather than days. That shows the argument isn't soritical. It doesn't rely on a day being a particularly short period of time.

21. It is also not the Epicurean position I'll discuss in the next section, since the lives between which Alex is choosing are equally long.

Future Dominance conflicts not only with the timeless goods view, but also with more modest non-separable views. The conflict arises when a good that one will enjoy tomorrow contributes to lifetime well-being in a way determined by context after tomorrow, or vice versa. Consider, for example, a view on which pleasure makes a decreasing marginal contribution to lifetime well-being. On this view, the more episodes of a pleasure a life contains, the less each episode contributes to lifetime well-being. In order to avoid a sharp line between one's life as a whole and one's life starting at one day old, the proponent of this view should allow that pleasure makes a decreasing marginal contribution to any periods that contain it. Suppose that, in *Marriage*, Alex will enjoy some pleasure tomorrow whatever she does, and that divorcing will give her more pleasure after tomorrow than staying married. But suppose that staying married will give her more of some other good, like friendship or meaningfulness, that does not make a decreasing marginal contribution. Now it looks like divorcing could give her an equally good day tomorrow and a slightly better future after tomorrow, but a slightly worse future as a whole, due to pleasure's decreasing marginal contribution to her future as a whole. So even more modest non-separable views should be rejected as inconsistent with *Future Dominance*.

My second argument appeals to a different principle concerning the future. The principle is:

Prudential Harmony If one is in a position to choose between life *x* and life *y*, and *x* offers a better future than *y*, then *x* offers a better life than *y*.

Before I defend the principle, let me explain why non-separable views violate it. They violate it in cases in which a future event or good makes a contribution to lifetime well-being that is determined by context in the past. In *Marriage*, skip ahead a day, to the point where Alex is deciding whether to fix her marriage or to divorce. Divorcing will give her a better future. But, on the timeless goods view, fixing the marriage could give her a better life as a whole. Or consider a view on which pleasure makes a decreasing marginal contribution to lifetime well-being. Suppose someone who has had a solitary but pleasant past is now choosing between a future richer in pleasure and one richer in friendship. Even if the former is the better future, it could offer the worse life as a whole due to pleasure's decreasing marginal contribution. The same would go if friendship makes an increasing marginal contribution (e.g., on the weak lexical superiority view), or if balance is important. All these non-separable views violate *Prudential Harmony*.

But views that violate *Prudential Harmony* face a trilemma. Suppose we have to choose between the better future and the better life as a whole. What

should we choose? It looks like there are three possibilities: (1) we should choose the better life as a whole, (2) we should choose the better future, or (3) we are permitted to choose either the better life as a whole or the better future. I'll consider them in turn.[22]

(1) Suppose we ought to choose the better life as a whole. The problem with this proposal is that it makes prudence systematically conflict with our pattern of self-interested preferences. As Derek Parfit observes, we prefer pain in the past to pain in the future. We would rather learn that we had a painful surgery yesterday than learn that we will have a painful surgery tomorrow. Moreover, we tend to think this preference rationally permissible. Noticing that we are biased towards the future does not occasion self-criticism in the way that, say, learning that we have cyclic preferences or even a bias towards the nearer future might.[23]

We have future-biased preferences regarding pain and pleasure. Intuitions may be less clear for non-hedonic goods, but it is hard to think of a good that elicits definite intuitions to the contrary—to the effect that we are indifferent to whether the good is past or future. So we generally seem to prefer the better future to the better life. This is in tension with option (1), which holds that prudence—what we should do *for our own sake*—requires us to choose the better life over the better future. Admittedly, our preference for pain to be in the past is practically inert: unless time travel is possible, we cannot ask the surgeon to go back in time and perform the painful operation

22. A similar argument can be found in Ben Bradley, "Narrativity, Freedom, and Redeeming the Past," *Social Theory and Practice* 37, no. 1 (2011): 47–62, https://doi.org/10.5840/soctheorpract20113714, 59–61. I build on Bradley's argument in several ways: by discussing future bias, by showing that option (2) makes prudence directly self-defeating, and by considering option (3).

23. Parfit, *Reasons and Persons*, chapter 64. Some argue that future-biased preferences are irrational: see, for example, Tom Dougherty, "On Whether to Prefer Pain to Pass," *Ethics* 121, no. 3 (2011): 521–37, https://doi.org/10.1086/658896; Tom Dougherty, "Future-Bias and Practical Reason," *Philosophers' Imprint* 15, no. 30 (2015): 1–16, http://hdl.handle.net/2027/spo.3521354.0015.030; Preston Greene and Meghan Sullivan, "Against Time Bias," *Ethics* 125, no. 4 (2015): 947–70, https://doi.org/10.1086/680910; and Dale Dorsey, *A Theory of Prudence* (Oxford University Press, 2021), https://doi.org/10.1093/oso/9780198823759.001.0001. For some responses, see Greene and Sullivan, "Against Time Bias," 953–56; David Braddon-Mitchell, Andrew Latham, and Kristie Miller, "Can We Turn People into Pain Pumps? On the Rationality of Future Bias and Strong Risk Aversion," *Journal of Moral Philosophy* 21, nos. 5–6 (2023): 593–624, https://doi.org/10.1163/17455243-20234084; and Dorsey, *A Theory of Prudence*, chapter 11. For defenses of future bias, see Caspar Hare, "A Puzzle about Other-Directed Time-Bias," *Australasian Journal of Philosophy* 86, no. 2 (2008): 269–77, https://doi.org/10.1080/00048400801886348 and Todd Karhu, "What Justifies Our Bias Toward the Future?" *Australasian Journal of Philosophy* 101, no. 4 (2023): 876–89, https://doi.org/10.1080/00048402.2022.2047747. This is a live debate, but, given the ubiquity and intuitive appeal of future-biased preferences, the claim that they're irrational faces an uphill battle.

yesterday.[24] So option (1) will not require us to act against our preferences, at least in cases where what is at stake is pain or pleasure. But it is in tension with what our preferences seem to show: that what we care about, for our own sake, is our future.[25]

(2) Suppose, then, that we ought to choose the better future. This has the strange consequence that the good life is prudentially irrelevant, except to newborns. But it has a stranger consequence: it compels us to make inconsistent decisions over time. When Alex marries, say at age 30, she should, insofar as she can, commit to fixing her marriage if it ever runs into trouble. For example, she should, if she can, sign a prenuptial agreement disincentivizing divorce. That is because fixing the marriage when it runs into trouble will give her the best future from 30. But, when her marriage does run into trouble, say when she's 50, concern for her future should lead her to divorce. If she had signed the prenup at 30, she should try to tear it up at 50. Or imagine you're Alex's parent. When Alex is young, you should advise her that, if she ever faces a scenario like *Marriage*, she should repair her marriage. When the time comes, however, you should advise her to divorce. It is implausible that prudence should require this sort of dynamic inconsistency.[26]

The worry can be sharpened. The view under consideration makes prudence *directly self-defeating*.[27] One can be faced with a sequence of choices such that, if one always does what is prudentially required, one will be worse off, relative to each time, than if one always does what is prudentially forbidden. Here is an example. (Although my argument does not entail that well-being can be cardinally represented, I will suppose for the sake of illustration here that it can be. The cardinal values can be straightforwardly replaced with ordinal rankings.)

Snowboard & Marriage At age 30, Alex can decide to learn to snowboard. Learning to snowboard will involve two decades of unpleasant effort—she's

24. Though Dougherty, "On Whether to Prefer Pain to Pass" shows these preferences can have practical relevance if one is risk-averse.

25. One might wonder whether we prefer the better future to the better life in cases where our preferences are relevant to action. Suppose Alex can choose the more meaningful, and thus better, life by saving her marriage, or the more meaningful, and thus better, future by divorcing. Is it intuitive that she would prefer the better future? The problem with this question is that it supposes Alex can be in a position to choose between the better life and the better future—an assumption which I am disputing.

26. Some have thought dynamic inconsistency rationally permissible in other cases. But it would be especially bad in this case, because it would be required rather than merely permitted, and it wouldn't depend on idiosyncratic (e.g., intransitive, nearness-biased, risk-averse, or incommensurable) preferences. It would be a deep feature of prudence, applying to all of us.

27. The term comes from Parfit, *Reasons and Persons*, 53.

not a natural—with a moderate payoff in enjoyment later. Learning to snowboard will subtract 2 units of well-being from her future from 30 (through the end of her life), but add 3 units of well-being to her future from 50. At age 50, Alex can get divorced. Divorcing will subtract 3 units of well-being from her future from 30 (due to its narrative disvalue), but add 1 unit of well-being to her future from 50.

Table 1 describes Alex's decision problem.

	Learn to snowboard	*Don't learn to snowboard*
Fix marriage	⟨–2,3⟩	⟨0,0⟩
Divorce	⟨–5,4⟩	⟨–3,1⟩

Table 1. Alex's choices. Payoff format: ⟨future from 30, future from 50⟩

At age 30, Alex prudentially ought not to learn to snowboard: not learning to snowboard will give her the better future regardless of what she chooses at age 50. At age 50, Alex prudentially ought to divorce: divorcing will give her the better future regardless of what she chose at age 30. If she does what she prudentially ought, Alex will get the outcome ⟨–3,1⟩: –3 from age 30 and 1 from age 50. But, if she learns to snowboard and fixes her marriage, she'll get ⟨–2,3⟩: a better future from age 30 and a better future from age 50. So, if she always does what she prudentially ought not to do, she will get a better outcome relative to each time. This is implausible.[28]

One might wonder whether this dynamic inconsistency is all that bad, since future-biased preferences already give rise to a sort of dynamic inconsistency. In the case alluded to earlier, Parfit imagines one is to have either a seriously painful surgery on Tuesday or a mildly painful surgery on Thursday. The following pattern of preferences seems rational: on Monday, prefer the latter; on Wednesday, prefer the former.[29] But the sort of dynamic

28. Samuel Fullhart writes: "Across a wide range of debates in ethics, decision theory, political philosophy, and formal epistemology, philosophers have assumed that if a normative theory is [directly] self-defeating, then that fact alone shows that the theory is defective." See Samuel Fullhart, "Embracing Self-Defeat in Normative Theory," *Philosophy and Phenomenological Research* 109, no. 1 (2024): 204–25, https://doi.org/10.1111/phpr.13033, 204. Fullhart contends there is no knockdown argument against directly self-defeating theories in general. I am not sure whether direct self-defeat is a problem in all the settings Fullhart mentions, but it does seem to me that the present instance of it, involving choices made by just one person, is counterintuitive.

29. I thank an anonymous referee for pressing me here.

inconsistency I am considering here is different in three respects. First, it concerns prudence rather than rational preference. Plausibly, prudence is less permissive than rational preference: we are rationally permitted to have imprudent preferences (think, for example, of altruistic preferences), but also rationally permitted to have prudent ones. Second, it is a matter of requirement rather than mere permission. Third, it concerns action rather than preference. On Wednesday, one cannot move the surgery back from Thursday to Tuesday. Each difference makes the dynamic inconsistency I am considering more objectionable than that which attends future-biased preferences. The combined effect of all three differences would make prudence directly self-defeating.

(3) Suppose, then, that we are permitted to choose the better life as whole or the better future.[30] This view might be motivated by a view about rational preference: it might be thought rationally permissible to *prefer* either the better life or the better future.[31] But prudence is not so permissive. If we are prudentially permitted to choose the better life and prudentially permitted to choose the better future, then there is nothing prudentially *special* about the future: it is of no more fundamental importance than the past. But, since we are prudentially permitted to ignore our well-being in the past, we must also be prudentially permitted to ignore our well-being in the future. So, for example, we must be prudentially permitted to choose the better past. And this implies that we must be prudentially permitted to sign up for torture, since torture in the future will not make us worse-off in the past. But this is absurd.

None of (1), (2), or (3) looks attractive. This trilemma gives us reason to endorse *Prudential Harmony*, reject non-separable views, and accept *Separability*.[32]

30. In "Narrativity, Freedom, and Redeeming the Past," Bradley considers a different proposal: perhaps there are two *varieties* of prudence, one concerned with the future, and one concerned with life as a whole. Bradley objects that this makes apparently meaningful questions like "what should I do?" meaningless (or, perhaps, ambiguous). Another drawback of this position is that, for reasons parallel to those I will give momentarily, it seems hard to avoid a proliferation of varieties of prudence, including one concerned only with the past.

31. See Samuel Scheffler, "Temporal Neutrality and the Bias Toward the Future," in *Principles and Persons: The Legacy of Derek Parfit*, eds. Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan (Oxford University Press, 2021), 85–114, https://doi.org/10.1093/oso/9780192893994.003.0005.

32. To be clear, *Separability* is consistent with some of what might be called "shape" views. For example, it's consistent with the view that it's better for daily well-being to be steady over the course of one's life rather than variable, or vice versa. That's because this view does not require a day's contribution to lifetime well-being to depend on its context. Suppose life *x* is always moderate while life *y* contains highs and lows. *Separability* says that which is better is determined

It is worth noting that this second argument for *Separability* is logically more ambitious than the first. It rules out not only non-separable views, but also the devolution view discussed above. Suppose, as the devolution view holds, that by saving her marriage Alex can improve her past days by making them more meaningful. In virtue of improving her past days, she can improve her life as a whole—without improving her future. Then saving the marriage could give her the better life as a whole, while divorcing might still give her the better future. This violates *Prudential Harmony*.

To recap the argument of this section: *Separability* is intuitively appealing, and non-separable views have two sorts of implausible implications in cases in which we think prudentially about the future. So *Separability* is true.

III. Longer Life

My second premise is:

Longer Life Extending a life by a very drab day leaves lifetime well-being unchanged.

Longer Life can be motivated by a case.

Treatment Bryce is eighty years old and seriously ill. We can let him die tomorrow, or we can give him a treatment that will extend his life by a month. The extra month will be decent for him, though not as good as his earlier life. He is temporarily unresponsive, so we cannot ask him what he wants.

Here is an argument: (i) Insofar as we care about Bryce, we should give him the treatment. (Imagine he is someone you care about.) But it is a truism that (ii) insofar as we care about Bryce, we should do what is best for him.[33] Therefore, (iii) it is best for him that we give him the treatment.

---

by the days on which they're different—that is, by whether it's better to have moderate days or highs and lows. *Separability* is also consistent with the view that it's better for daily well-being to slope upwards rather than downwards (see §*IV*). Suppose life *z* starts low and ends high while life *w* starts high and ends low. *Separability* says that their relative ranking is determined by well-being on the days in which they differ: the early days and the late days. It's consistent with *Separability* that *z* is better because it is worse in the early days and better in the late days—that is, because it is upward-sloping. See also Broome, *Weighing Lives*, 225–28.

33. Stephen Darwall takes this sort of claim to be definitional of what is good for us: see Stephen Darwall, *Welfare and Rational Care* (Princeton University Press, 2002), https://doi.org/10.1515/9781400825325. It seems to me a truism even if not definitional. Perhaps it should be restricted to cases in which doing what's best for Bryce doesn't violate any moral side-constraint, for example by killing him or by treating him as a mere means. But letting him die here doesn't do any such thing.

It is a short step from (iii) to *Longer Life*. We can reduce the quality of the extra month down to the point at which we should be indifferent about whether to give Bryce the treatment. A month of very drab days seems to do the trick. At that point, giving him the treatment will be just as good for him as withholding it. And varying the details of Bryce's earlier life does not affect the truth of (i). Were we to find out that his childhood was happier than we thought, or that he was older than we thought, that should not affect what we choose to do.

Here is another way of understanding *Longer Life*. In ordinary speech, to say that someone's life is worth living is to say that it is better for that person to continue to live than to die.[34] *Longer Life* says that, if tomorrow is to be one's last day and it is to be very drab, then one's life is now just on the border between worth living and not worth living.

A few philosophers reject *Longer Life*. They think that, in a case like *Treatment*, it could be better for Bryce to die early. David Velleman says an early death can make for a better life story. Thomas Hurka proposes, by comparison, that a shorter career can be a better career, and therefore a greater achievement. For example, he writes, Muhammad Ali's career would have been better had it not included the last, mediocre fights against Larry Holmes and Trevor Berbick, even though these fights were decent in themselves.[35]

If these philosophers want to reject (iii), should they reject (i) or (ii)? In fact, they seem to reject (i). They seem to think that, if Bryce has had a wonderful life so far, we should deny him a decent extra month at the end of it.[36]

But this makes a cruel fetish of life stories. It is one thing for Ali's friends to hope he retires. It is another for them to hope he dies. (In fact, even if the fights against Holmes and Berbick gave Ali a worse career, I doubt they diminished his *achievement*.)

Moreover, people do not generally seem to be moved by the sorts of reasons Velleman and Hurka think they have to die early. Evidence for this comes from surveys of people who seek voluntary assisted dying (VAD) in

34. Here I follow John Broome's analysis of "worth living": see John Broome, *Weighing Goods: Equality, Uncertainty and Time* (Basil Blackwell, 1991), 167.

35. See Velleman, "Well-Being and Time," 62 and Thomas Hurka, *Perfectionism* (Oxford University Press, 1993), 71, https://doi.org/10.1093/0195101162.001.0001. There is some evidence that non-philosophers share this judgment: see Ed Diener, Derrick Wirtz, and Shigehiro Oishi, "End Effects of Rated Life Quality: The James Dean Effect," *Psychological Science* 12, no. 2 (2001): 124–28, https://doi.org/10.1111/1467-9280.00321.

36. Velleman writes that "a person may rationally be willing to die even though he can look forward to a few more good weeks or months" (Velleman, "Well-Being and Time," 62). Hurka takes the Muhammad Ali case to imply we should sometimes end our lives early (Hurka, *Perfectionism*, 74).

jurisdictions where VAD is legal. A study in the Netherlands reports that patients' most common reasons for seeking VAD are pointless suffering (67%), deterioration or loss of dignity (65%), and weakness or tiredness (56%).[37] More telling, however, is that the study's authors only offer reasons to do with patients' current or future situation. They even ask explicitly whether patients' reasons pertain mainly to their current situation or to their future situation—with no mention of their past situation or their life as a whole. In Western Australia, a government survey reports that patients' most common reasons for seeking VAD are loss of ability to engage in enjoyable activities (65%), loss of autonomy (65%), and loss of dignity (53%).[38] Again, more telling is that no reasons are offered concerning patients' past situation or their life as a whole. Both surveys do offer an "other" option—which would encompass life-story reasons—but few patients report "other" reasons (5% in the Netherlands, 6% in Western Australia).

One might object to my interpretation of these studies. Law in the Netherlands and Western Australia requires that a patient seeking VAD be suffering from a disease that is hopeless (the Netherlands) or terminal (Western Australia). So we should expect patients seeking VAD to have reasons related to suffering. But this does not preclude them from having additional reasons. And indeed they do, including worries about dignity (noted above) and about being a burden to family (18% in the Netherlands, 35% in Western Australia). My interpretation is also bolstered by evidence from Switzerland. Although Switzerland does not require suffering or illness as a condition of eligibility for VAD, 98.5% of those who died by VAD in the 2010–2014 period reported a concomitant illness.[39]

Some might still choose to die early despite having a good future to look forward to. One reason for this might be the risk of losing control: someone who does not choose to die while she is still mentally competent to do so might find herself locked into an undesired end. But it is conceivable that one could choose to die early for the reasons Velleman and Hurka posit. We

37. Marijke Jansen-van der Weide, Bregje Onwuteaka-Philipsen, and Gerrit van der Wal, "Granted, Undecided, Withdrawn, and Refused Requests for Euthanasia and Physician-Assisted Suicide," *Archives of Internal Medicine* 165, no. 15 (2005): 1698–704, https://doi.org/10.1001/archinte.165.15.1698. The study surveys physicians regarding 1,681 patients who requested VAD.

38. Western Australia, Voluntary Assisted Dying Board, *Annual Report, 2023–2024* (2003), https://www.health.wa.gov.au/~/media/Corp/Documents/Health-for/Voluntary-assisted-dying/VAD-Board-Annual-Report-2023-24.pdf. The study surveys 525 patients deemed eligible for VAD.

39. Swiss Confederation, Federal Statistical Office, *Cause of Death Statistics 2014: Assisted Suicide and Suicide in Switzerland* (2017), https://www.bfs.admin.ch/bfs/en/home/statistics/catalogues-databases/publications.assetdetail.3902308.html.

might not, in practice, be inclined to criticize such a person. People making difficult end-of-life choices deserve respect and compassion. But if anyone is likely to have a life story, achievement, etc. that would be improved by an early death, many VAD-eligible patients should. We should doubt Velleman's and Hurka's view about prudential reasons to die early because people who seem well-positioned to have these reasons do not appear to be moved by them.

Perhaps Velleman and Hurka should instead reject (ii). Perhaps they should say that, insofar as we care about Bryce, we should do what is best for his future, not for his life as a whole. (Perhaps the truism that we ought to do what is best for him is ambiguous on this point.) But this position violates *Prudential Harmony*. As I noted in §II, doing so leads to dynamic inconsistency. On this view, when my son is born, I should hope that, if he is in Bryce's situation, he will not get the treatment. But, when he is eighty and in Bryce's situation, I should (if I am still alive) hope he will get the treatment. The view is also, as I showed, directly self-defeating.

That said, I think we should accept *Longer Life* even if we are open to rejecting *Prudential Harmony*. Imagine being a young person expecting to live eighty wonderful years. Would you want to live another twenty merely decent years afterwards, even though they would represent a decline? I suspect many would, and would want this for their loved ones. Moreover, when I reflect on the question, it seems to me that what matters is not how good these twenty years would be relative to my first eighty, but how good they would be in themselves. Would I be incapacitated? Would my friends still be alive? Would I regret that I could no longer do what I loved? If these are the questions that matter, and if drab days are just barely good enough in themselves, then *Longer Life* follows.

Epicurus offers a different argument against *Longer Life*. Death, he writes, "is relevant neither to the living nor to the dead, since it does not affect the former, and the latter do not exist."[40] John Broome reconstructs Epicurus's argument as follows. (iv) If one life is worse than another for Bryce, then it is worse for him at some time. (v) A shorter life is not worse for Bryce at any time, because it is not worse to be dead than to be alive. Therefore, (not-iii) it is not better for Bryce that we give him the treatment.[41]

40. Epicurus, *Epistle to Menoeceus* 125, in *The Epicurus Reader: Selected Writings and Testimonia*, eds. Brad Inwood and Lloyd Gerson (Hackett, 1994), 29.

41. John Broome, *Ethics out of Economics* (Cambridge University Press, 1999), 170–73. See also Broome, *Weighing Lives*, 235–40. David Hershenov interprets Epicurus differently: as interested in the badness of death, rather than the relative value of different lives. See David Hershenov, "A More Palatable Epicureanism," *American Philosophical Quarterly* 44, no. 2 (2007): 171–80.

Epicurus's view is extreme. It entails that a life of twenty decent years followed by eighty wonderful years is no better than a life of twenty decent years followed by an early death. Most take this to be a *reductio* of the view.[42] Epicurus's argument is interesting because it forces us to reject either (iv) or (v), but I do not here need to decide which.

I will conclude this section by introducing some terminology. Let a *level* of daily well-being be a class of days equal in well-being. Let Γ_0 be the level of a very drab day, and let Γ_1 be the level of a slightly better but still drab day. It follows from *Longer Life* that an extra day at Γ_0 is just on the border between worth living and not worth living. An extra day at Γ_1, then, is barely worth living.[43] When I prove the *Unhappy Conclusion*, I will use Γ_1 as the level of well-being in every day of the drab life *z*.

IV. Later Isn't Worse

My third premise I take to be uncontroversial. It concerns the timing of good days within a life. Some philosophers say it does not matter whether good days come earlier or later. They endorse:

Time-Invariance If two lives are such that, for every level of daily well-being, both lives offer the same number of days at that level, then the two lives are equally good.[44]

Others say it is better for well-being to trend upwards over the course of one's life. This is the *shape-of-a-life hypothesis*.[45] These philosophers say a life is better if good days occur later rather than earlier. They do not typically opine on the value of shapes that are not flat or monotonically increasing or monotonically decreasing. But I take it they do not think a life is made better by a good day's occurring earlier rather than later—that is, by having more of a downward trend. So I take it they would join friends of *Time-Invariance* in accepting:

---

42. For a survey, see Travis Timmerman, "Dissolving Death's Time-of-Harm Problem," *Australasian Journal of Philosophy* 100, no. 2 (2022): 405–18, https://doi.org/10.1080/00048402.2021.1891108.

43. Though it would be question-begging to say that a life with every day at Γ_1 "would always be barely worth living" (Parfit, "Overpopulation and the Quality of Life," 18). If the *Unhappy Conclusion* is true, then a sufficiently long life at Γ_1 would, in the beginning, be very much worth living.

44. See, for example, Henry Sidgwick, *The Methods of Ethics*, 7th ed. (Macmillan, 1907), 381 and Feldman, *Pleasure and the Good Life*, chapter 6.

45. See, for example, Michael Slote, *Goods and Virtues* (Oxford University Press, 1983) and Joshua Glasgow, "The Shape of a Life and the Value of Loss and Gain," *Philosophical Studies* 162, no. 3 (2013): 665–82, https://doi.org/10.1007/s11098-011-9788-0.

Later Isn't Worse Starting from one life (here construed as an assignment of well-being levels to days), if a better day is moved later and a worse day earlier, the resulting life is at least as good as the starting life.[46]

In other words, a more upward-sloping (or less downward-sloping) life is at least as good as a less upward-sloping (or more downward-sloping) one. This is my third premise.

One might, I suppose, disagree on the grounds that it is better for good days to come in the middle of life rather than in youth or old age.[47] The proof of the *Unhappy Conclusion* can be modified to accommodate this or any other more nuanced shape view consistent with *Separability*, as long as it is not much worse for good days to occur later.

V. Tradeoffs

My fourth premise concerns tradeoffs between different levels of daily well-being within a life of fixed length. To state the premise, I will define one more term. Say that one level of daily well-being Γ_i is *connected up* to a higher level Γ_j just in case some life with every day at Γ_i is at least as good as some equally long life with one day at Γ_j and every other day at Γ_0. The intuitive idea is that a lifetime at Γ_i is worth a day at Γ_j. Now, the premise is:

Tradeoffs Between Γ_1 and any higher level of daily well-being, there is a finite sequence of levels such that each level is connected up to the next.[48]

In defending *Tradeoffs*, I will focus on the sorts of good days we can easily imagine. I will show that, for these sorts of days, the sequence contemplated by *Tradeoffs* will be short—perhaps just two or three levels.[49]

46. A formal statement may be clearer: For any two lives x and y of equal length and days t and u within them such that u is later than t, if $x_u \sim y_t \succ x_t \sim y_u$ and, for all days v other than t and u, $x_v \sim y_v$, then $x \succcurlyeq y$. When I say the premise would be widely accepted, I mean at least conditional on *Separability*.

47. This is one way of reading the view in Slote, *Goods and Virtues*.

48. This premise is loosely inspired by the Archimedean and solvability axioms in David Krantz, Duncan Luce, Patrick Suppes, and Amos Tversky, *Foundations* of *Measurement*, vol. 1 (Academic Press, 1971), 253–56. To be clear, it is consistent with the view that there's a limit to how good a single day can be.

49. This should allay any worry that my argument here relies on a sorites series, as other arguments against the lexical superiority view have been alleged to do. See Griffin, *Well-Being*, 87; Teruji Thomas, "Some Possibilities in Population Axiology," *Mind* 127, no. 507 (2018): 807–32, https://doi.org/10.1093/mind/fzx047; Teruji Thomas, "Are Spectrum Arguments Defused by Vagueness?" *Australasian Journal of Philosophy* 100, no. 4 (2022): 743–57, https://doi.org/10.1080/00048402.2021.1920622; and Nebel, "Totalism Without Repugnance," 217–21.

Here's an example to illustrate *Tradeoffs* and show why it is plausible. Imagine you are to live a hundred-year life with every day at Γ_1. You will enjoy only a very mild good on each day—perhaps a sitcom, or an afternoon nap, or, in Parfit's phrase, muzak and potatoes.[50] Now suppose you can give up all these goods, lowering every day to Γ_0, in exchange for just one very good day on your fiftieth birthday. Let Γ_2 be the level of well-being on your fiftieth birthday that would make you indifferent. How good would Γ_2 have to be to compensate for the loss of a lifetime of sitcoms, afternoon naps, or muzak and potatoes? I encourage you to think of your own answer. But, conservatively, I assume it must include something at least as good as a meal at a fine restaurant or a long vigorous workout.

Now repeat the exercise. Imagine you are to live a hundred years with every day at Γ_2. You have the option to lower every day to Γ_0 in exchange for a better day on your fiftieth birthday. Let Γ_3 be the level of well-being on your fiftieth birthday that would make you indifferent. How good would Γ_3 have to be to compensate for the loss of a lifetime of fine meals or long vigorous workouts? Again, I encourage you to think of your own answer. For myself, I doubt that even a good day of my own life would do the trick. If you disagree, however, repeat the exercise again to Γ_4, and to Γ_5. I suggest, in short order, you will reach as good a day as can easily be imagined.[51] That is what *Tradeoffs* says. And going through the example shows why it is plausible. (If, at some step, no day is good enough to do the trick, that is no problem: *Tradeoffs* only says that, at each step, a lifetime at the lower level is as good *or better* than one day at the higher level and every other day at Γ_0.)

Should the proponent of the lexical superiority view about well-being reject *Tradeoffs*? It is not obvious that she should. *Tradeoffs* is logically more modest than the negation of the lexical superiority view. In effect, *Tradeoffs* posits a sequence of levels of well-being, with each level such that a lifetime of it is at least as good as a day of the next. The lexical superiority view only denies that a lifetime of the *first* is at least as good as a day of the *last*. In fact, most supporters of the lexical superiority view seem to favor what I called the weak version of the view, which only denies that a lifetime of the first is at least as good as some minimum amount—perhaps a decade—of the last. *Tradeoffs* is thus logically more modest than the negation of the lexical superiority view, and doubly more modest than the negation of the weak version of that view. It is also intuitively easier to accept *Tradeoffs* than to deny the

50. Parfit, "Overpopulation and the Quality of Life," 18.

51. Anecdotally: the handful of people whom I have asked to do this exercise have reported that a single step is enough.

lexical superiority view: we need only make several pairwise comparisons between lives of equal, normal length, rather than trying to evaluate arbitrarily large amounts (millions of years' worth?) of very mild pleasures.

VI. The Unhappy Conclusion

I will now prove that *Separability*, *Later Isn't Worse*, *Longer Life*, and *Tradeoffs* together entail the *Unhappy Conclusion*. For simplicity, I will start by assuming *Time-Invariance*. The idea behind the proof is that we start with a happy life, add a number of days at Γ_0 on at the end, then take a series of steps that decrease well-being in earlier days while increasing it in later days. Each step leaves lifetime well-being at least as high, and the process yields a life with every day at Γ_1. The proof is a bit technical, and the reader who wishes may safely skip to the next section.

First, a lemma. Suppose Γ_i is connected up to Γ_j. That means that there is a life x with every day at Γ_i that's at least as good as some equally long life y with one day at Γ_j and every other day at Γ_0. Say that x and y are n days long and that y offers Γ_j on day t. The lemma is that, if we start with any life containing at least one day at Γ_j and at least $n-1$ days at Γ_0, decrease one day from Γ_j to Γ_i, and increase $n-1$ days from Γ_0 to Γ_i, the resulting life will be at least as good as the life with which we started. In effect, the lemma shows we can take the sort of tradeoff contemplated by *Tradeoffs* and embed it within a longer life. *Separability*, *Longer Life*, and *Time-Invariance* together guarantee this.[52]

Now to the *Unhappy Conclusion*. Start with any life *a*. Let Γ_{high} be the level of well-being on the first day of *a*. By *Tradeoffs*, there is a finite sequence $\Gamma_1, \ldots, \Gamma_{medium}, \Gamma_{high}$ in which each level is connected up to the next. Let n be the length of a pair of lives that witnesses the connection between Γ_{medium} and Γ_{high}. Now perform the following two-step procedure. First, add $n-1$ days to Γ_0 to the end of *a*. The resulting life is just as good as *a* (by *Longer*

52. Proof:

[1] $x \succcurlyeq y$.
[2] (Any life comprising n days at Γ_i) $\succcurlyeq$ (Any life comprising n days, all at Γ_o except for day t at Γ_j) (*Separability*).
[3] (Any life comprising n days at Γ_i followed by m days at Γ_o) $\succcurlyeq$ (Any life comprising n days, all at Γ_o except for day t at Γ_j, followed by m days at Γ_o) (*Longer Life*).
[4] (Any life comprising n days at Γ_i and m days at Γ_o) $\succcurlyeq$ (Any life comprising $(n+m-1)$ days at Γ_o and one day at Γ_j) (*Time-Invariance*).
[5] (Any life comprising n days at Γ_i and m days at any other levels) $\succcurlyeq$ (Any life comprising $n-1$ days at Γ_o, one day at Γ_j, and m days at the same levels as the same m days on the left-hand side) (*Separability*).

Life). Second, decrease well-being on the first day from Γ_{high} to Γ_{medium}, and increase well-being on the $n-1$ newly added days from Γ_0 to Γ_{medium}. The resulting life is at least as good as the starting life (by the lemma), and therefore at least as good as *a*. Now repeat the procedure on the resulting life, adding some days at Γ_0 at the end, decreasing well-being on the first day from Γ_{medium} to the preceding level in the sequence, and increasing well-being on the newly added days accordingly. A finite number of repetitions of this procedure will bring the first day down to Γ_1. Now repeat the procedure for the rest of the days that have well-being above Γ_1, including days in the original life and days added later. (If any day of *a* has well-being below Γ_1, simply increase that day's well-being to Γ_1.) The lemma ensures that each iteration of the procedure yields a life that is at least as good as the starting life. The result of all the iterations will be a long life with every day at Γ_1 that is at least as good as *a*. Now add a single day at Γ_0 to the end and increase that day to Γ_1. The resulting life *z* has every day at Γ_1 and is better than *a*. That proves the *Unhappy Conclusion*.

Each iteration of the procedure just described involved decreasing well-being on an earlier day while increasing well-being on some later days. If *Time-Invariance* is false but *Later Isn't Worse* is still true, then daily well-being later in life counts for more. Each iteration will have a more positive effect on lifetime well-being than if *Time-Invariance* is true. So each iteration will still yield a life that is at least as good as the starting life, and the *Unhappy Conclusion* will still be true.

VII. Connection to Population Ethics

The *Unhappy Conclusion* is an intrapersonal analog of the *Repugnant Conclusion*. But the present argument for the *Unhappy Conclusion* is not merely an analog of any existing argument for the *Repugnant Conclusion*. The one which it most closely resembles is Parfit's "mere addition" argument.[53] Parfit appeals to the premise that "merely adding" a person with a drab life to a population makes the population no worse. *Longer Life* is more or less analogous to this premise. But Parfit also appeals to the premise that equalizing well-being across people while slightly increasing the average makes a population better. While this non-anti-egalitarian premise is intuitively appealing,

53. Parfit, *Reasons and Persons*, 419–30. See also Parfit, "Overpopulation and the Quality of Life," 11–14. Variations on the argument include the second impossibility theorem in Gustaf Arrhenius, "Future Generations: A Challenge for Moral Theory" (PhD diss., Uppsala University, 2000) and the "up-down" argument in Derek Parfit, "Can We Avoid the Repugnant Conclusion?" *Theoria* 82, no. 2 (2016): 110–27, https://doi.org/10.1111/theo.12097.

its analog in the setting of well-being over time is less so. The life of highs and lows has some appeal. That is the central idea of the lexical superiority view. And that is why I do not appeal to the analog of Parfit's premise, but instead put pressure on the lexical superiority view through my other premises.[54]

That said, given the analogy between the *Unhappy Conclusion* and the *Repugnant Conclusion*, one might hope to find a way to resist the former by canvassing strategies for resisting the latter. There are, broadly speaking, three strategies for resisting the *Repugnant Conclusion*.[55] The first is to reject mere additions. The second is to adopt the lexical superiority view about general good. The third is to reject transitivity. I will show that neither the first nor the second strategy yields a promising, dialectically effective response to the present argument for the *Unhappy Conclusion*. The third strategy, rejecting transitivity, seems to me to work equally well in both cases, if it works at all.

There are several ways to resist mere additions. According to the *averagist* and *variable-value views*, merely adding a person can make a population worse when it decreases average well-being.[56] According to the *critical-level view*, there is some level of well-being that is itself quite good, but is such that adding even a good life below that level makes a population worse.[57] The analog of any of these views in the setting of well-being over time would imply that, in a case like *Treatment*, it can be worse for Bryce to live the extra month, even if the month itself would be good. But this is deeply counterintuitive. And our intuitions about cases like *Treatment* are informed by experience. We normally think about well-being when deciding whether to extend people's lives. We do not normally think about population ethics when deciding, say, whether to have children.

54. I thank an anonymous referee for encouraging me to explain why the arguments are structurally disanalogous.

55. For an overview, see Gustaf Arrhenius, Jesper Ryberg, and Torbjörn Tännsjö, "The Repugnant Conclusion," in *The Stanford Encyclopedia of* Philosophy, Winter 2022 Edition, eds. Edward Zalta and Uri Nodelman, https://plato.stanford.edu/archives/win2022/entries/repugnant-conclusion/.

56. See Thomas Hurka, "Value and Population Size," *Ethics* 93, no. 3 (1983): 504–5, https://doi.org/10.1086/292462. A different variable-value view is proposed by Theodore Sider, "Might Theory X Be a Theory of Diminishing Marginal Value?" *Analysis* 51, no. 4 (1991): 265–71, https://doi.org/10.1093/analys/51.4.265, but this view is non-separable.

57. See Charles Blackorby, Walter Bossert, and David Donaldson, "Critical-Level Utilitarianism and the Population-Ethics Dilemma," *Economics and Philosophy* 13, no. 2 (1997): 197–230, https://doi.org/10.1017/s026626710000448x; Charles Blackorby, Walter Bossert, and David Donaldson, *Population Issues in Social Choice Theory, Welfare Economics, and Ethics* (Cambridge University Press, 2005), chapter 5, https://doi.org/10.1017/CCOL0521825512; and Broome, *Weighing Lives*.

The *person-affecting view* offers a different way of resisting mere additions. According to this view, one population is better than another only if it is better *for* someone.[58] Adding even a very happy person does not make a population better. The proponent of the view can hold that adding a very happy person leaves general good unchanged. But then general good will be the same no matter how happy that person is, which is implausible. More promisingly, the proponent of the view can hold that any two populations comprising different people are incommensurable in value.[59] The analog of the person-affecting view in the setting of well-being over time is the view that one life is better than another only if it is better *at* some time. This is just the Epicurean view. As I said earlier, the view is extreme—more extreme, I think, than the person-affecting view has seemed to its proponents to be. It may be tempting to think that two populations are incommensurable when they contain different people. But it is not tempting to think that two lives are incommensurable when they contain different (numbers of) days.

In sum, *Longer Life* is more secure than its population-ethics analog. And that makes intuitive sense. There is less daylight, so to speak, between well-being in a day and well-being in a lifetime than there is between well-being in a lifetime and general good in a population.

The second strategy for resisting the *Repugnant Conclusion* is to adopt the lexical superiority view about general good. According to this view, there are two kinds of goods, such that some amount of a higher good is better, in general (that is, for a population), than an arbitrarily large amount of a lower good. The idea is that a happy population will contain higher goods, whereas a drab population will not.

Since this view concerns the general good of a population, it is different than the lexical superiority view about well-being. But one might motivate it by appealing to the lexical superiority view about well-being. Parfit does so—and in turn motivates the lexical superiority view about well-being by assuming that the *Unhappy Conclusion* is false.[60] One might (as Parfit appears

58. See Jan Narveson, "Utilitarianism and New Generations," *Mind* 76, no. 301 (1967): 62–72, https://doi.org/10.1093/mind/lxxvi.301.62; Parfit, *Reasons and Persons*, chapter 16; David Heyd, "Procreation and Value: Can Ethics Deal with Futurity Problems?" *Philosophia* 18, nos. 2–3 (1988): 151–70, https://doi.org/10.1007/bf02380074; and Ralf Bader, "Person-Affecting Utilitarianism," in *The Oxford Handbook of Population Ethics*, eds. Gustaf Arrhenius, Krister Bykvist, Tim Campbell, and Elizabeth Finneron-Burns (Oxford University Press, 2022), https://doi.org/10.1093/oxfordhb/9780190907686.013.20.

59. This point is made by John Broome, "Should We Value Population?" *Journal of* Political *Philosophy* 13, no. 4 (2005): 399–413, https://doi.org/10.1111/j.1467-9760.2005.00230.x and Bader, "Person-Affecting Utilitarianism."

60. Parfit, "Overpopulation and the Quality of Life," 17–20. Griffin, *Well-Being*, chapter 5 and Portmore, "Does the Total Principle Have Any Repugnant Implications?," 84–87 argue similarly.

to do) argue by analogy from the lexical superiority view about well-being to the lexical superiority view about general good. Or one might argue from *totalism*, the claim that the general good of a population corresponds to the total well-being of everyone in it.[61] Either way, the present argument for the *Unhappy Conclusion* threatens to undermine this strategy. That is because, as I noted earlier, it more or less follows from the *Unhappy Conclusion* that the lexical superiority view about well-being is false.

One might be inclined to accept the lexical superiority view about general good not because one already accepts the lexical superiority view about well-being, but simply because one wants to resist the *Repugnant Conclusion*. If so, then one could, without circularity, argue from the lexical superiority view about general good to the lexical superiority view about well-being. But that will not reveal which premise of the present argument for the *Unhappy Conclusion* is false.[62]

I will close by presenting an argument from the *Unhappy Conclusion* to the *Repugnant Conclusion*. C. I. Lewis and R. M. Hare offer the following criterion for ranking populations: the general good of a population corresponds to the well-being one would enjoy by living each of its lives in sequence.[63] If the Lewis-Hare criterion is true, then the *Unhappy Conclusion* entails the *Repugnant Conclusion*. Given any happy population *A*, let *a*

Portmore, for example, writes that the lexical superiority view about general good "is supported by the fact that we do seem to prefer a certain amount of life that is well worth living to any amount of life that is so drab as to be only barely worth living," ibid., 84. The lexical superiority view about general good is popular: see Philip Kitcher, "Parfit's Puzzle," *Noûs* 34, no. 4 (2000): 550–77, https://doi.org/10.1111/0029-4624.00278; Thomas, "Some Possibilities in Population Axiology;" Nikhil Venkatesh, "Repugnance and Perfection," *Philosophy and Public Affairs* 48, no. 3 (2020): 262–84, https://doi.org/10.1111/papa.12165; Erik Carlson, "On Some Impossibility Theorems in Population Ethics," in *The Oxford Handbook of Population Ethics*, eds. Gustaf Arrhenius, Krister Bykvist, Tim Campbell, and Elizabeth Finneron-Burns (New York: Oxford University Press, 2022), https://doi.org/10.1093/oxfordhb/9780190907686.013.14; and Nebel, "Totalism Without Repugnance."

61. Nebel, "Totalism Without Repugnance," offers a version of totalism with lexical superiority.

62. Alternatively, one might hold that the higher goods make a lexical contribution to general good but not to well-being. Simon Beard takes this to be Parfit's view: see Simon Beard, "Perfectionism and the Repugnant Conclusion," *The Journal of Value Inquiry* 54, no. 1 (2020): 119–40, https://doi.org/10.1007/s10790-019-09687-4, 122–23. But in some places Parfit does endorse the lexical superiority view about well-being: see, for example, Parfit, "Overpopulation and the Quality of Life," 19.

63. C. I. Lewis, *An Analysis of Knowledge and Valuation* (Open Court, 1946), 546–47 and R. M. Hare, *Moral Thinking: Its Levels, Method, and Point* (Oxford University Press, 1981), 129, https://doi.org/10.1093/0198246609.001.0001. The view is endorsed by Cowen, "Normative Population Theory," 34–35 and Portmore, "Does the Total Principle Have Any Repugnant Implications?" 85–86n12. It is also discussed sympathetically by Roger Crisp, "Utilitarianism and the Life of Virtue," 150–51.

be the life obtained by living each life in *A* in some sequence. The *Unhappy Conclusion* entails that there is a life *z* containing only drab days that is better than *a*. Let *Z* be a population containing lives of normal length that, if lived in some sequence, would constitute *z*. It follows from the Lewis-Hare criterion that *Z* is better than *A*, which proves the *Repugnant Conclusion*.

Is the criterion true? It can be defended on the ground that it regiments an intuitive idea: the value of a population is its value *for everyone*. More modestly, it reflects the idea that the value of a population depends on the lives of the people who comprise it, and that everyone counts equally. The criterion can also be motivated by higher-level theoretical considerations. It permits a theoretical economy by allowing us to do without two distinct value notions of general good and well-being.[64] It also permits us to avoid aggregating well-being. This will be a theoretical benefit if well-being is not measurable, and so not aggregable. It will also be a theoretical benefit if the aggregation of well-being is somehow conceptually defective. It might be thought that aggregating well-being can only yield, well, aggregate well-being, but that there is no such thing, since well-being must be *someone's*. (Though I note these higher-level considerations, I will not take a position on them, as I do not think we need to appeal to them to motivate the criterion's extensional correctness.)

A natural worry about the criterion is that it falls prey to John Rawls's objection to utilitarianism: that it "does not take seriously the distinction between persons."[65] Rawls's objection can be understood in several ways.[66] One is as the claim that a harm to one person cannot always be compensated by an equally large benefit to another. But, as Tyler Cowen notes, the Lewis-Hare criterion does not entail this, because it does not assume that a harm in one part of a life can be compensated by an equally large benefit in another.[67] Indeed, the criterion is consistent with various views about distributional goods, including egalitarianism. It permits us to motivate these views by appealing to views about well-being. This gives it *ad hominem* force against those proponents of the lexical superiority view about general good, noted

64. Indeed, it may permit us to do without either. Lewis and Hare think that, to decide between two outcomes, we need only vividly imagine living each life in sequence and decide which we would prefer. The criterion may thus appeal to those who hope to naturalize value by reducing it to preference.

65. John Rawls, *A Theory of Justice* (Harvard University Press, 1971), 27, https://doi.org/10.2307/j.ctvjf9z6v

66. For some that are not relevant here, see Richard Yetter Chappell, "Value Receptacles," *Noûs* 49, no. 2 (2015): 322–32, https://doi.org/10.1111/nous.12023.

67. Cowen, "Normative Population Theory," 42.

above, who seek to motivate their view by appealing to the lexical superiority view about well-being.

Another way of understanding Rawls's objection is as a conceptual worry. Rawls might be worried that it is conceptually problematic to sum or otherwise aggregate well-being across people. But, as I noted above, it is, if anything, a virtue of the Lewis-Hare criterion that it escapes this objection.[68]

Although my main quarry in this paper has been the *Unhappy Conclusion*, the Lewis-Hare criterion permits an appealing argument from that to the *Repugnant Conclusion*.

VIII. Conclusion

I argued for the *Unhappy Conclusion* from four premises. The first was *Separability*, which I defended by bridging the gap between daily well-being and lifetime well-being with future well-being. The second was *Longer Life*, a premise whose appeal derives from ordinary judgments about when we ought to prolong someone's life. The third was *Later Isn't Worse*, which I took to be uncontroversial. The fourth was *Tradeoffs*, which I defended with a thought-experiment about tradeoffs between pairs of lives of normal length. These premises entail the *Unhappy Conclusion*. Since the present argument avoids the analogs of two popular strategies for resisting the *Repugnant Conclusion*, the *Unhappy Conclusion* is on surer footing than its population-ethical cousin.

Acknowledgements Thanks to Ryan Doody, Caspar Hare, Adam Pautz, Bernard Reginster, Asher Shang, several anonymous referees, and especially Jamie Dreier. Thanks also to participants in the 2024 Lund Value Additivity Workshop and an audience at the 2025 APA Eastern Division meeting.

Competing Interests The author has no competing interests to declare.

68. Another possible objection: it is impossible for one person to live many lives in sequence, because the physical and psychological changes between the lives (or "lives") would be so drastic as to render their subjects distinct people. But this objection appeals to the wrong sort of possibility. It might be metaphysically impossible to live many lives in sequence, but it is not conceptually impossible. Many people believe in the possibility of reincarnation, and they are not conceptually confused. As long as we can conceive of living all the lives in sequence, we can make true value judgments about doing so.