# MINIMALIST AXIOLOGIES

Alternatives to 'Good Minus Bad' Views of Value

TEO AJANTAIVAL

EDITED BY MAGNUS VINDING

Preface

Can suffering be counterbalanced by the creation of other things?

Our answer to this question depends on how we think about the notion of positive value.

In this book, I explore ethical views that reject the idea of *intrinsic* positive value, and which instead understand positive value in relational terms. Previously, these views have been called purely negative or purely suffering-focused views, and they often have roots in Buddhist or Epicurean philosophy. As a broad category of views, I call them minimalist views. The term "minimalist axiologies" specifically refers to minimalist views of value: views that essentially say "the less this, the better". Overall, I aim to highlight how these views are compatible with sensible and nuanced notions of positive value, wellbeing, and lives worth living.

A key point throughout the book is that many of our seemingly intrinsic positive values can be considered valuable thanks to their helpful roles for reducing problems such as involuntary suffering. Thus, minimalist views are more compatible with our everyday intuitions about positive value than is usually recognized.

This book is a collection of six essays that have previously been published online. Each of the essays is a standalone piece, and they can be read in any order depending on the reader's interests. So if you are interested in a specific topic, it makes sense to just read one or two essays, or even to just skim the book for new points or references. At the same time, the six essays all complement each other, and together they provide a more cohesive picture.

Since I wanted to keep the essays readable as standalone pieces, the book includes significant repetition of key points and definitions between chapters. Additionally, many core points are repeated even within the same chapters. This is partly because in my 13 years of following discussions on these topics, I have found

that those key points are often missed and rarely pieced together. Thus, it seems useful to highlight how the core points and pieces relate to each other, so that we can better see these views in a more complete way.

I will admit upfront that the book is not for everyone. The style is often concise, intended to quickly cover a lot of ground at a high level. To fill the gaps, the book is densely referenced with footnotes that point to further reading. The content is oriented toward people who have some existing interest in topics such as philosophy of wellbeing, normative ethics, or value theory. As such, the book may not be a suitable first introduction to these fields, but it can complement existing introductions.

I should also clarify that my focus is broader than just a defense of my own views. I present a wide range of minimalist views, not just the views that I endorse most strongly. This is partly because many of the main points I make apply to minimalist views in general, and partly because I wish to convey the diversity of minimalist views.

Thus, the book is perhaps better seen as an introduction to and defense of minimalist views more broadly, and not necessarily a defense of any specific minimalist view. My own current view is a consequentialist, welfarist, and experience-focused view, with a priority to the prevention of unbearable suffering. Yet there are many minimalist views that do not accept any of these stances, as will be illustrated in the book. Again, what unites all these views is their rejection of the idea of intrinsic positive value whose creation could by itself counterbalance suffering elsewhere.

The book does not seek to present any novel theory of wellbeing, morality, or value. However, I believe that the book offers many new angles from which minimalist views can be approached in productive ways. My hope is that it will catalyze further reflection on fundamental values, help people understand minimalist views better, and perhaps even help resolve some of the deep conflicts that we may experience between seemingly opposed values.

All of the essays are a result of my work for the Center for Reducing Suffering (CRS), a nonprofit organization devoted to

reducing suffering. The essays have benefited from the close attention of my editor and CRS colleague Magnus Vinding, to whom I also directly owe a dozen of the paragraphs in the book. I am also grateful to the donors of CRS who made this work possible.

Contents

Part I: Varieties of Minimalist Views

Part II: Minimalist Views in Theory and Practice

Part III: Positive Roles in the Real World

Part I

Varieties of Minimalist Views

Chapter 1

Minimalist Views of Wellbeing

1.1 Introduction

1.1.1 Problems with "Good Minus Bad" Views

What is wellbeing? Hedonistic views typically define it as pleasure minus pain. Preference-based views might define it as the difference between our fulfilled and unfulfilled preferences. Finally, objective list views might posit a number of independent goods and bads that each contribute to our overall wellbeing.

These "good minus bad" views of wellbeing all face a shared challenge, as they all rely on an offsetting theory of aggregation. That is, they assume that any independent bads can always be counterbalanced or offset by a sufficient addition of independent goods, at least within the same life. (Consider Figure 1.1.)

This offsetting premise has more problems than are commonly recognized, including the often sidelined question of what

justifies it in the first place.[1] In population ethics, it plays a key role in generating moral implications such as 'Creating Hell to Please the Blissful'.[2] At the individual level, the offsetting premise implies that a rollercoaster life that contains arbitrary durations of unbearable agony and a sufficient amount of non-relieving goods (Figure 1.2) has greater wellbeing than does a completely un-troubled life. These issues highlight the importance of exploring alternative conceptions of wellbeing that do not rely on the offsetting premise.

Figure 1.1. All else equal, can agony always be offset by adding enough happy moments elsewhere, even within a life?

---

[1] Vinding, 2020c, 2022e.

[2] More in Chapter 4.
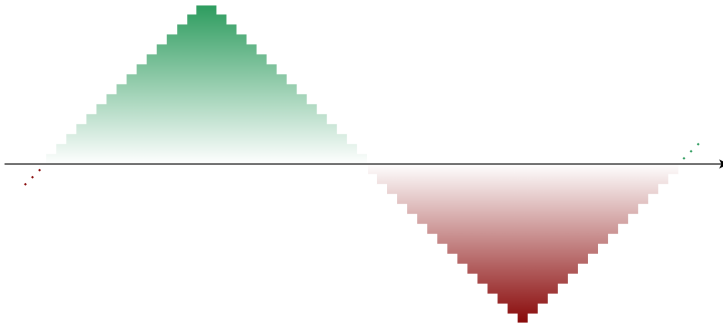
Figure 1.2. The supposedly positive balance of a rollercoaster life of unbearable agony plus a sufficient amount of non-relieving goods.

1.1.2 Minimalist Alternatives

Minimalist views work without the concept of independent goods. Instead, they define things that make our life go better for us in entirely relational terms, via the absence of sources of illbeing. In other words, minimalist views take a fully relational approach to positive value, and thereby avoid the problems that are specific to the offsetting premise.[3]

Minimalist views are often overlooked in existing introductions to wellbeing theories, which tend to focus only on the variety of "good minus bad" views on offer.[4] Yet not only do minimalist

---

[3] I use the term 'relational value' as a synonym for 'instrumental value' to avoid the latter term's misleading connotations. These connotations might lead us to perceive this value narrowly or unwisely as "merely instrumental… merely a tool", or as somehow diminished relative to the actual importance of the value in question. More in 6.2.

[4] For instance, minimalist views of wellbeing are overlooked in the introductions by Fletcher (2016); the Stanford Encyclopedia of Philosophy (Crisp, 2021); the Happier Lives Institute (Moorhouse et al., 2020); and Utilitarianism.net (Chappell & Meissner, 2023). But minimalist views of wellbeing have been defended in the philosophical literature, such as by Schopenhauer (1819, "the negative nature of all satisfaction"; 1851, "negative in its character"; see also Fox, 2022, "pleasures of distraction"); Fehige (1998); Breyer (2015); Sherman (2017); and Knutsson (2021b, sec. 3, "about negative well-being").

views deserve serious consideration for their comparative merits, they can also be positively intuitive in their own right.[5]

In particular, minimalist views can make sense of the practical tradeoffs that many of us reflectively endorse, with no need for the offsetting premise in the first place. And because many minimalist views focus on a single common currency of value, they are promising candidates for resolving conflicts between multiple, seemingly intrinsic values. By contrast, all "good minus bad" views are still pluralistic in that they involve at least two distinct value entities.[6]

Although minimalist views do not employ the concept of an independent good, they still provide principled answers to the question of what makes life better for an individual. And in practice we would be wise to consider the narrow question of what is 'better for oneself' within the broader context of what is 'better overall'. In this context, all non-egoistic minimalist views agree that life can be worth living and protecting for its overall positive roles.[7]

In this chapter, I briefly explore the variety of minimalist views of wellbeing, not to provide an exhaustive survey, but to give a sense of their diversity and intuitive appeal.

For instance, *experientialist* minimalist views, such as tranquilism, align with the "experience requirement", which is the intuition that our wellbeing cannot be directly affected by things outside our experience.[8] In contrast, *extra-experientialist* mini-

---

[5] More on the comparative implications of minimalist and offsetting views in Chapter 4 and Chapter 5.

[6] For some reasons that might motivate a preference for value monism, see the debate between monists and pluralists on the topic of value incommensurability, Schroeder, 2018, sec. 2.2.3.

[7] More in Chapter 3 and Chapter 6.

[8] On experientialism about wellbeing, see van der Deijl, 2021.

malist views, such as antifrustrationism or objective list minimal-
ism, reject the experience requirement.[9] Thus, they can align with
the intuition that premature death can leave us worse off, such as
when it results in frustrated preferences, violation of autonomy, or
ruined life projects.[10]

The next section will outline some of the main problems with
the offsetting premise of "good minus bad" views. The rest will
explore what might be the most intuitive ways to think of well-
being without it.

1.2 Reasons to Doubt the Offsetting Premise: A Brief Overview

The offsetting premise posits that any independent bads can al-
ways be counterbalanced or offset by a sufficient addition of in-
dependent goods. To motivate the exploration of minimalist views
of wellbeing, this section will briefly outline some of the main
reasons to doubt the offsetting premise.[11]

1. **Lack of comprehensive defenses**:[12] Major defenses of
   offsetting views lack comprehensive defenses of the off-
   setting premise, even though this premise has been re-
   jected in various ways and faces additional challenges

---

[9] Not all objective list views reject the experience requirement, but the
ones I examine do, hence the categorization used in this brief chapter.

[10] In other words, the extra-experientialist minimalist views in this chap-
ter agree that we can have reasons to stay alive even for our own sake,
namely to live a life that is better for us than a life cut short. And at the
normative level, all minimalist views can support practical reasons to
strongly oppose cutting lives short (5.2.6.2). These reasons include the
preservation of relationally positive roles, lives, and norms.

[11] The points in this section are summarized mostly from Chapter 4,
Chapter 5, and Vinding, 2020c, 2022e. Some of the points may overlap
with each other; they are numbered primarily for readability.

[12] Vinding, 2020c, sec. 1.

compared to minimalist views. (These various rejections and challenges are outlined below.)

2. **Problems with existing defenses**:[13] Existing defenses of the offsetting premise often rely on thought experiments intended to show that alternative views would have implausible implications.[14] Yet these defenses tend to be unconvincing for a number of reasons. Specifically, they …

   a. … are confounded by various other issues and practical intuitions that are not supposed to influence our judgment in the relevant cases.[15]

   b. … often assume the existence of "higher" states above a subjectively flawless or completely undisturbed state to begin with. Yet a strong case can be made that a completely undisturbed state is rarely if ever reached during our waking hours, and that a completely undisturbed state is plausibly the hedonic ceiling.[16]

   c. … rarely address extra-experientialist minimalist views that avoid the purportedly implausible

---

[13] Vinding, 2020c, sec. 6.

[14] See Vinding, 2020c, sec. 6.1, sec. 6.2; 2022b; 2022f; 2022h.

[15] Such confounders are discussed at least in 4.3.4; 5.1.1; 5.2.4; 5.2.6; Knutsson, 2015; and Vinding, 2022k. These confounders also include misleading framings that are prone to paint a picture that is far from what a minimalist would endorse as an accurate understanding of their views (Vinding, 2020c, sec. 6.1; 2022f, sec. 11, sec. 15). For instance, beings may be described as experiencing states of "mere neutrality" — of being "reduced" or "not particularly happy" — where the accurate understanding would be "completely untroubled by any experiential bads" (cf. Knutsson, 2022b; Gloor, 2017, sec. 4.2).

[16] This argument is made in Knutsson, 2022b. I convey a sense of the main idea in Appendix 1.

implications without relying on the offsetting premise.[17]

    d.   … rarely attempt to show that the implications of minimalist views would be worse than those of offsetting views. A 'side-by-side' comparison provides reason to think that the opposite is the case: that the minimalist implications are the least repugnant.[18]

3. **Various views reject the premise**:[19] The offsetting premise has been rejected in various ways, such as by the views below. (The first two will not be explored further in this chapter.)

    a.   **Incommensurate value entities**:[20] Some philosophers hold that pleasure has intrinsic value, but deny that pleasure can compensate for suffering.[21]

    b.   **Lexical priority views centered on severe bads**:[22] Lexical views grant categorical priority to certain value entities relative to others. Such views are often centered on severe bads.[23] For instance, many people have the intuition that torture-level suffering cannot be counterbalanced by any purported good.[24]

---

[17] Vinding, 2022a, sec. 2.1.

[18] 4.4; 5.2.5.

[19] Vinding, 2020c, sec. 2.

[20] Vinding, 2020c, sec. 2.1.

[21] Wolf, 1997. Compare also the 'xNU+' framework of Leighton, 2023.

[22] Vinding, 2020c, sec. 2.5.

[23] Knutsson, 2016c, "Lexicality of bads over goods".

[24] Tomasik, 2015c, "Pain-pleasure tradeoff"; Gloor, 2016, "Torture-level suffering cannot be counterbalanced".

    c. **Minimalist views**:[25] Minimalist views entirely reject the notion of independent goods, and instead understand good in relational terms.[26] Closely associated is the normative view that ethics is about solving problems and not about creating unneeded goods that do not relieve anyone's burden ("non-relieving goods").[27]

4. **Additional challenges compared to minimalist views**:[28] It is widely accepted that severe bads can outweigh milder ones, such as when we prioritize severe suffering over minor pains in triage situations. Yet this notion of "negative outweighing", shared by both offsetting and minimalist views, is fundamentally different from the additional offsetting premise, according to which problematic states like severe suffering could be "canceled out" or "made up for" by the creation of unneeded goods elsewhere. The offsetting premise thus faces the additional challenges of how to justify …

    a. … that subjectively unbearable agony, or any bads at all, could be counterbalanced by non-relieving goods (as well as what these goods are, and why some of them are better than others).[29]

    b. … why a rollercoaster life of "unbearable agony plus a sufficient amount of non-relieving goods" (Figure 1.2) has greater wellbeing than a completely untroubled life.

---

[25] Vinding, 2020c, sec. 2.3.

[26] More on the relational view of goods in Chapter 3 and Chapter 6.

[27] For more on the view that ethics is about problems, see 3.3.4; Vinding, 2020c, sec. 2.4; and Leighton, 2023, chap. 4, "The Notion of Urgency".

[28] Vinding, 2020c, sec. 4, sec. 5.

[29] Vinding, 2020c, sec. 3.2.

c. … why offsetting views would be more plausible in the context of population ethics, where a strong case can be made that they share all the most "repugnant" features of minimalist views while introducing additional sources of repugnance.[30]

5. **A priori reasons to doubt phenomenological dual opposites**:[31] Offsetting views are sometimes defended with reference to a phenomenologically opposite counterpart to suffering. Yet the idea of such dual opposite dimensions of experience is less parsimonious and less simple than a view without such opposites.[32] Additionally, it is unclear whether the notion of phenomenological opposites even makes sense, and what it would be like for two experiential states to be dual opposites.

---

[30] This case is made in Chapter 4. A counterexample might be replacement arguments (discussed in Chapter 5 and Knutsson, 2021b), where purely consequentialist versions of experientialist minimalist views would imply that there is nothing suboptimal about replacing any world with an empty world ("cessation"). Yet how repugnant are such cessation implications if we already assume a purely consequentialist and experience-focused view? When we properly account for the consequentialist equivalence between cessation and non-creation (5.2.2), this question is equivalent to the question of how repugnant it is to not create any set of experiences for their own sake. And here one may find non-creation to be wholly non-repugnant (4.3.4). By contrast, the replacement implications of consequentialist offsetting views seem worse, including the implication to replace completely untroubled lives with rollercoaster lives (cf. Chapter 4; 5.2.5).
(Other minimalist views, explored in 1.3.2, may consider cessation suboptimal due to extra-experiential bads such as preference frustration.)

[31] Vinding, 2022e, sec. 3.1.

[32] Specifically, the assumption of dual opposites implies a more bloated and less sparse ontology, and goes against old ideals of avoiding unnecessary postulations in our explanatory models. (Thanks to Simon Knutsson for comments on this point.)

6. **Lack of introspective evidence**:[33] A direct argument against the existence of a phenomenological counterpart to suffering is that, for many of us, introspection yields no sign of such a counterpart. When we introspectively examine various candidates of positive experiences, we seem to find no phenomenological properties that would render them dual opposites of suffering.[34] (Readers are encouraged not to take this claim on authority, but to earnestly pursue this introspective exercise themselves.)

7. **Deflationary or debunking explanations for why we might believe in a positive counterpart**:[35] There are reasonable alternative explanations for the belief in a positive counterpart to suffering:

   a. **A misprojection of our common tendency to think in terms of positive and negative real numbers**:[36] We tend to think in terms of real numbers and basic addition, since they offer a powerful conceptual framework that is perfectly valid in many contexts. Consequently, we might project these numbers onto our experiences, even if introspection or other evidence might ultimately fail to support such a conceptual representation in this domain.[37]

---

[33] Vinding, 2022e, sec. 3.2.

[34] Vinding, 2022e, sec. 1.

[35] Vinding, 2022e, sec. 4.

[36] Vinding, 2022c, sec. 4.

[37] And more broadly, our styles of thinking may be influenced by a deep cultural history where it often made practical sense to quickly divide various things into opposite categories, such as 'good–bad', 'positive–negative', and 'right–wrong'. But this way of dichotomous, dualistic, or absolutizing thinking is not the only way of thinking, and while it might make sense in mathematical domains, it might be better seen as a practical heuristic elsewhere. We might not even gravitate there ourselves, had we grown up thinking comparatively in terms of betterness relations. See

b. **Confusion between relative and absolute "pluses"**:[38] We might mistake a reduction in discomfort, boredom, or the like (a relative "plus") for a genuinely positive experience (an absolute "plus"), when in fact the latter may not exist.[39]

1.3 The Variety of Minimalist Views

1.3.1 Experientialist Views

Experientialist views say that our wellbeing cannot be directly affected by things outside our experience.[40] Experientialist minimalist views define wellbeing as the degree to which we are free from experiential sources of illbeing, such as pain, suffering, disturbance, or a visceral non-acceptance of our current experience. Such views often draw inspiration from the Buddhist or Epicurean philosophical traditions.

At first glance, the various kinds of experientialist minimalist views may seem to differ in only semantic or aesthetic ways. And

---

also Knutsson, 2023a, "Scepticism of categorical value notions", "Thinking comparatively".

[38] Vinding, 2022e, sec. 4.

[39] See Sherman (2017, p. 8, sec. 11.2), Gloor (2017, sec. 2.1), and Knutsson (2022b, sec. 5.2). Additionally, Fox (2022) explains how Schopenhauer maintained that all pleasures feel only relatively good, because they are either (1) "pleasures of satisfaction": satisfying a prior lack or need, or (2) "pleasures of distraction": reducing our awareness of a lack or need. Both types of pleasure strongly correlate with a relative freedom from felt dissatisfaction, which could explain why we might mistake them for inherently positive experiences instead of just relatively positive changes compared with our preceding states. (A third type of "relatively positive experience" could be the prediction-based "anticipation of positive future changes", namely changes that are desirable for preventing some dissatisfaction that we intuitively care about, even if this dissatisfaction is not currently ours; cf. Gloor, 2017, sec. 4.5; Vinding, 2022e, sec. 7.)

[40] van der Deijl, 2021.

perhaps they are more similar to each other than are extra-experientialist views. Yet they still differ substantially in how they understand the nature of experiential wellbeing.

For instance, some views might say that the unpleasant quality of an experience is intrinsically bad for us in an objective or 'attitude-independent' way, regardless of our own stance on it. By contrast, other views are more subjective or 'attitude-sensitive'. That is, they might equate wellbeing with things like how much we ourselves wish for our experience to change: the less, the better. Tranquilism, in particular, posits that no experience is inherently desirable nor undesirable, and focuses instead on our subjective need for change.[41]

1.3.1.1 Buddhist Minimalism

Some minimalist views of wellbeing are based directly on Buddhist ideas. For instance, according to philosopher and scholar of Buddhism Daniel Breyer, the Pāli Buddhist tradition understands wellbeing as consisting in the cessation of *dukkha* (suffering, 'disease', dissatisfaction), with other factors being good for us only insofar as they contribute to the cessation of *dukkha*.[42]

1.3.1.2 Epicurean Minimalism

Similar minimalist views may be inspired by the Epicurean tradition. For instance, a kind of minimalist hedonism results from rejecting the concept of positive pleasure at the descriptive level, and adopting instead these Epicurean-inspired concepts of kinetic and static pleasures:[43]

1. Kinetic pleasure: "Kinetic pleasure is the active removal of a pain".[44] It is "what results from fulfilling a desire" or

---

[41] Gloor, 2017, sec. 5.

[42] Breyer, 2015.

[43] These ideas are traced to their sources and explored much further in Sherman, 2017, and Knutsson, 2019, 2022b.

[44] Sherman, 2017, p. 53.

when a "lack or need is being removed".[45] It is what we usually call pleasure, yet it is actually not an independent good but rather a temporary relief from a prior experiential disturbance.

2. Static (*katastematic*) pleasure: Static pleasure is the upper limit that the kinetic, remedial pleasure leads up to, "a condition of absolute contentment in mind and body".[46] There is no higher pleasure: "The limit of pleasure is reached with the removal of all pain"; freedom from all pain is the summit of pleasure.[47]

Together, these concepts are taken to cover all experiences of pleasure. For instance, "joy is active, the removal of mental pain, while tranquillity is the static state of being without [any] distress".[48] Thus, there are ultimately just more or less disturbed experiential states, and according to Epicurean-inspired minimalist views, greater wellbeing consists in reducing our experiential disturbances. (For more on this, see Appendix 1.)

1.3.1.3 Contraction-Based Minimalism

In various interviews and conversations, meditator Roger Thisdell has outlined a minimalist view of wellbeing that is rooted in his own phenomenological observation. The view is formulated in the following terms.[49]

1. Contraction versus pleasure: All experiences contain some level of unpleasantness or disturbance, which Thisdell calls 'contraction'. This contraction exists at a

---

[45] Cf. Knutsson, 2019.

[46] Sherman, 2017, p. 106.

[47] Cf. Knutsson, 2022b.

[48] Sherman, 2017, p. 53.

[49] Gómez-Emilsson, 2021; Perry, 2022; Prest & Thisdell, 2022; Knutsson & Thisdell, 2023.

more concrete, phenomenologically pinpointable level than does pleasure. Pleasure is, on closer inspection, more like a comparative judgment or evaluation that occurs after experiencing a relief from contractive pressure.[50]

2. Betterness and cessation: As one feels less and less contraction (and thus more expansion), one feels better and better. Yet, the complete absence of disturbance is only achieved in moments when all experiences and phenomenology cease. According to Thisdell, this undisturbed state is what meditators call 'cessation', which is beginning to be studied in neuroscience.[51]

1.3.1.4 Tranquilism

Tranquilism as proposed by Lukas Gloor is a view of wellbeing inspired by Buddhist and Epicurean ideas. It understands wellbeing in the following way.[52]

1. Wellbeing and cravings: Wellbeing is the degree to which we are free from cravings, which are defined as need-based, visceral desires to change something about our current experience. Cravings are what make an experience bad for us. For instance, the sensation of pain, without any

---

[50] Cf. the Epicurean notion of kinetic pleasure, 1.3.1.2.

[51] Laukkonen et al., 2023. Regarding how common Thisdell's view of wellbeing is, Thisdell thinks that many people familiar with deep meditation might agree that there is no higher state than a completely peaceful one, but that many such people might not use an analytical framework that would lead them to compare the value of individual states in this isolated manner (Knutsson & Thisdell, 2023).

[52] Tranquilism by itself is an experientialist minimalist theory of wellbeing. Yet the author also notes in many places that it is only intended to be a theory of momentary experiential wellbeing and not a standalone moral theory, emphasizing how it is compatible with views that incorporate non-experiential aims.

craving for it to end, is not intrinsically bad according to Gloor.[53]

2. Optimal states: Optimal states are free from aversive components and from cravings for more pleasure. They include states of contentment such as meditative tranquility and flow states, as well as all subjectively flawless states of inner peace, dreamless sleep, and non-consciousness.[54]

3. Pleasure and cravings: Pleasure, understood as a sensation that increases our hedonic level, can be valuable for its roles in preventing and protecting against the formation of cravings. But the absence of unneeded pleasure is entirely unproblematic. A pleasureless state of consciousness, if free from all cravings, is considered perfectly optimal and happy.[55]

In other words, tranquilism emphasizes our 'inside view' and considers all subjectively untroubled states as optimal because they are not experienced as suboptimal. By contrast, an offsetting hedonism entails the 'outside view' that more pleasure would always be better for us, even if we do not crave or desire it.[56]

---

[53] Cravings are a specific kind of desire, distinct from preferences and reflection-based desires. Preferences, according to Gloor, are abstract constructs that are present at all times. Desires, in contrast, are seen as preferences that have been "activated" or turned into conscious goals. Cravings are the need-based, visceral desires that we often cannot help but develop. (Gloor, 2017, sec. 2.2)

[54] Gloor, 2017, sec. 2.1, sec. 3.

[55] Gloor, 2017, sec. 2.1, sec. 4.1, sec. 4.2, sec. 5.

[56] This respect for 'inside' over 'outside' desires about experiences is also the focus of two brief defenses of tranquilism by Vinding (2020d, p. 49) and DiGiovanni (2021b).

1.3.2 Extra-Experientialist Views

1.3.2.1 Preferentialist Views

Offsetting preferentialism holds that (1) satisfied preferences are good for us, (2) frustrated preferences are bad for us, and (3) the former can offset the latter.[57]

In contrast, professor of philosophy Christoph Fehige has defended a view called antifrustrationism, which holds only the second premise:

> We don't do any good by creating satisfied extra preferences. What matters about preferences is not that they have a satisfied existence, but that they don't have a frustrated existence. … Maximizers of preference satisfaction should instead call themselves minimizers of preference frustration.[58]

A similar view has been expressed in the past by Peter Singer: "The creation of preferences which we then satisfy gains us nothing. We can think of the creation of the unsatisfied preferences as putting a debit in the moral ledger which satisfying them merely cancels out."[59]

The main difference between experientialist minimalist views and Fehige's view is that, according to the latter,

---

[57] Preference-based views are also called "desire theories", and philosophers tend to use the words "preference" and "desire" interchangeably. But I will speak of "preferentialist views" and "preferences", to maintain a clear contrast to *experientialist* desire-based views such as tranquilism.

[58] Fehige, 1998, p. 518. See en.wikipedia.org/wiki/Antifrustrationism. More on antifrustrationism in the context of population ethics in 3.3.

[59] It is worth noting that Singer (1980) wrote favorably of combining Preference Utilitarianism and Classical Utilitarianism. Yet Singer appears to have moved further toward Classical Utilitarianism in recent years (see e.g. Lazari-Radek & Singer, 2017, chap. 3).

> People need not be aware of their preferences; what counts is rather the attitude they *would* have towards something *if* they fully represented it.[60]

See also the explanation and defense of 'negative ideal preference utilitarianism' found in the Negative Utilitarianism FAQ.[61]

1.3.2.2 Conditional Interests Versus Teleological Goods

Closely related to preferentialist minimalist views are views centered on conditional interests, as defended by philosopher Johann Frick.[62] These views are structurally similar to Fehige's view, except they apply to interests rather than preferences.[63,64]

Frick's view gives us reason to be skeptical of the concept of intrinsic value, as it criticizes the "teleological view of wellbeing", which holds that wellbeing is something to be "promoted" in an unconditional sense.

In Frick's words (emphases mine):

> According to the **teleologist**, the appropriate response to what is good or valuable is to promote it, ensuring that as much of it exists as possible …
>
> Next, note that viewing some value *F* as **to be promoted** implies that there is no deep moral distinction between [1] increasing the degree to which *F* is realized

---

[60] Fehige, 1998, p. 509.

[61] Anonymous, 2015.

[62] Frick, 2014, 2020. Frick's thesis and paper have also inspired the conditional interest views that have been explored, on the Effective Altruism Forum, by Michael St. Jules (2019a, 2019b) and Lukas Gloor (2022).

[63] St. Jules, 2019a.

[64] To highlight their possible differences: It might be in our interest that we, for instance, acquire more skills, freedom, or access to pain relief, even if we have no preference for it. Or we might prefer to take certain health risks, or undergo certain hardships, that are not in our own best interest.

> amongst *existing* potential bearers of that value, and [2] creating *new* bearers of that value.[65]

In contrast, Frick argues that we never have unconditional reasons to bring about new instances of wellbeing as a teleological good.[66] What instead matters is individuals' degree of wellbeing *conditional* on their existence, that is, whether their interests are satisfied or violated *if* they exist.

Thus, by analogy to how Fehige's view was about the minimization of frustrated preferences, one can understand Frick's view to be about the minimization of violated interests. That is, we have reasons to satisfy interests so that they are not violated, but we have no reasons to create new satisfied interests, all else being equal.[67]

1.3.2.3 Objective List Views

Lastly, there are minimalist versions of objective list views. Objective list views of wellbeing typically claim that various objective goods contribute independently to our wellbeing, where these objective goods may include things like personal achievements, knowledge, and autonomy.

---

[65] Frick, 2020, pp. 63–64.

[66] By analogy, Frick finds it striking how the teleological promotion approach seems even more problematic for moral values *other than* wellbeing, such as "[justice,] liberty, equality, fairness, honesty, fidelity, loyalty, promise keeping, gratitude, charity, health, safety, etc.":

> None of these values appear even remotely plausible as candidates for "promotion" [in the teleological sense]. For instance, while we recognize strong moral reasons to make people free and equal, freedom and equality clearly do not require us to create new people so that they, too, may instantiate these values. [Frick, 2020, pp. 65–66.]

[67] As Michael St. Jules (2019a) put it: "We accomplish no good by creating and then satisfying an interest, all else equal, because interests give us reasons for their satisfaction, not for their existence or satisfaction over their nonexistence."

Minimalist versions of objective list views can retain largely the same list of objective goods, yet the important difference is that they construe these "goods" purely in terms of the absence of bads. As described by Vinding regarding the purported objective goods of autonomy and virtuous conduct:

> For example, rather than seeing autonomy as an objective good that can bring our wellbeing above some neutral level, the absence of autonomy is seen as an objective bad that detracts from our wellbeing, placing us below a neutral or unproblematic state of wellbeing; and having full autonomy can at most bring us to an untroubled or unproblematic level of wellbeing. … Rather than seeing virtue as an objective good that contributes positively to wellbeing, vice is seen as an objective bad that contributes negatively, and virtue may be understood as the mere absence of vice (cf. Kupfer, 2011; Knutsson, 2022a, sec. 4). And so on for any other purported objective good.[68]

The list of objective bads that matter independently to our wellbeing may include things like compromised health, false beliefs, ignorance, premature death, uncompleted life projects, being constrained, being discriminated against, being exploited, being manipulated, being subjected to violence, and so on.

Objective list views need not entail that such objective conditions are the *only* things that matter to our wellbeing, but merely that certain objective conditions *also* matter. Thus, one may think that our wellbeing consists in both our objective conditions and our experiential states.

---

[68] Vinding, 2023. Minimalist objective list views can also support the view that premature death is bad in many ways: "For not only may these views consider premature death to be bad because it entails many other objective bads (e.g. death would prevent us from completing our life projects), but these views may also see premature death itself as an objective bad."

This brings us to a more general point, namely that the individual views explored in this chapter need not be endorsed as standalone views, but can also be combined into a wide variety of minimalist hybrid theories of wellbeing. For example, one may endorse a minimalist hybrid view according to which our experiential states, preferences, conditional interests, and objective conditions all contribute independently to our wellbeing.

This gives a sense of the potential flexibility and variety of minimalist views of wellbeing, and a sense of how there are many reasonable alternatives to "good minus bad" views.

Appendix 1

Expanding the Epicurean Notion of Freedom From All Pain

At first, we may find implausible the Epicurean notion that the highest pleasure is freedom from all pain.[69] Yet we might find it more intuitive once we unpack the full meaning of "all pain". As expanded in Knutsson (2022b, sec. 2.1), a completely undisturbed state is entirely free from any bothersome instances of a long list of things, including:

- ache, agitation, agony, alienation, angst, anguish, annoyance, anxiety, boredom, compression, confusion, contempt, dejection, depression, desolation, despair, desperation, discomfort, discontentment, disgust, dislike, dismay, disorientation, dissatisfaction, distress, dread, embarrassment, enmity, ennui, envy, fear, frustration, gloominess, grief, guilt, hatred, heartbreak, horror, hopelessness, humiliation, hurting, impatience, indignation, insecurity, irritation, jealousy, loneliness, longing, loathing, loss, malaise, melancholia, nausea, nervousness,

---

[69] Cf. 1.3.1.2.

pain, panic, queasiness, regret, rejection, remorse, resentment, restlessness, sadness, shame, sorrow, stress, suffering, tension, terror, throb, tiredness, trouble, unease, unsafety, vexation, want, weariness, Weltschmerz, worry;

- feelings of being or having been betrayed, disliked, exploited, harmed, let down, neglected, treated badly, underappreciated, unloved, unwanted, or used;

- feelings of being or having been a burden on, bad for, or harmful to others;

- feelings of meaninglessness;

- feelings of effort, resistance, and struggle;

- feeling burdened, constrained, lack of control, overworked, stuck, threatened, unfortunate, unfree, unlucky, or weak;

- feeling damaged, decaying, declining, defective, ignorant, ill, incompetent, like a bad person, like a failure, like an impostor, low self-esteem, stupid, ugly (or that part of oneself is ugly), unclean, unhealthy, worthless, or of low worth.

A three-part partition of quality of life into positive, neutral, and negative

A two-part partition of quality of life into flawless and flawed
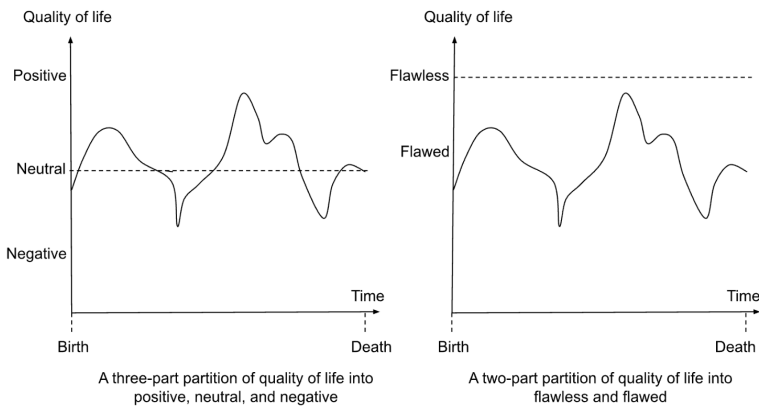
Figure A1.1. Two different ways to think of the quality of the same life over time. (Illustration by Simon Knutsson from Knutsson & Vinding, forthcoming.)

Chapter 2

Varieties of Minimalist Moral Views: Against Absurd Acts

2.1 Introduction

What are minimalist views?

- **Minimalist views of value (axiologies)** are evaluative views that define betterness solely in terms of the absence or reduction of independent bads. For instance, they might roughly say, "the less suffering, violence, and violation, the better". They reject the idea of weighing independent goods against these bads, as they deny that independent goods exist in the first place.

- **Minimalist moral views** are views about how to act and be that include a minimalist view of value, instead of an offsetting ('good minus bad') view of value.[70] They reject the concept of independently positive moral value, such

---

[70] Cf. 1.2.

as positive virtue or pleasure that could independently counterbalance bads.[71]

2.1.1 Alleged Recommendations of Absurd Acts

Overall, we may find minimalist views to be plausible alternatives to 'good minus bad' views. Yet minimalist views are sometimes alleged (at least in their purely consequentialist versions) to recommend absurd acts in practice, such as murdering individuals, or choosing not to save people's lives, so as to prevent their future suffering.[72]

My goal here is to broadly outline the various reasons why the most plausible and well-construed versions of these views — including their purely consequentialist versions — do not recommend such acts.

---

[71] Others may define "minimalist moral views" more broadly to also include views that reject independently positive or offsetting moral value without endorsing minimalist axiological claims of any kind (e.g. they may include views that lack an axiology for ranking different worlds in terms of betterness). Such views might include versions of fully nonconsequentialist views that entail only moral claims that are "minimalist in flavor", such as that we have moral reasons to reduce vice, harm, violations, and so on.

[72] For instance, an influential yet misleading essay by Toby Ord (2013) contains the following claim:

> [Negative utilitarianism, a consequentialist view focused on the minimization of suffering] implies that much healthcare and lifesaving is of enormous negative value. It says that the best healthcare system is typically the one that saves as few lives as possible, eliminating all the suffering at once. This turns healthcare policy debates on their heads and means we shouldn't be emulating France or Germany, but should instead look to copy failed states such as North Korea.

(This and other misleading claims in Ord's essay are given a point-by-point reply in Vinding, 2022f.)

For instance, in the case of purely consequentialist minimalist views, the consequentialist framework would be just as considerate of indirect, long-term effects as it would be in the offsetting versions of such views. This is worth noting because the purported absurd practical implications arguably don't stem from minimalism itself, but from its combination with implausible interpretations of pure consequentialism.

2.1.2 Only Straight Down in the Diagram: Minimalist Views Are Broader Than That

Minimalist views need not be purely consequentialist at the normative level. Similarly, purely consequentialist views need not be purely welfarist, and purely welfarist views need not be purely experience-focused. And even in the case of minimalist views that are purely experience-focused and purely consequentialist, one would still, in practice, give a lot of weight to many extra-experiential and seemingly nonconsequentialist considerations — such as the positive roles of autonomy, cooperation, and nonviolence — as part of a nuanced and impartial multi-level consequentialism (cf. Figure 2.1).

Varieties of minimalist moral views

Purely consequentialist? — No → **Nonconsequentialism**
"*Outcomes* aren't the only criterion of moral rightness."
Virtue ethics
Deontology
Social contract theory
Care ethics
Skepticism of general moral theories

Yes ↓

Purely welfarist? — No → **Extra-welfarist consequentialism**
"*Wellbeing* isn't the only morally relevant part of outcomes."
'Ideal consequentialism' and related views that give independent disvalue to certain acts.

Yes ↓

Purely experientialist? — No → **Extra-experientialist welfarism**
"*Experiences* aren't the only morally relevant part of wellbeing."
Views where wellbeing is affected also by preference frustrations, interest violations, or objective conditions.

Yes ↓

Experientialist welfarism

Is consequentialism used as a decision procedure?

Not necessarily ↗ **Rule consequentialism**
"*Practical decisions* follow established rules that, if generally applied, would yield the best overall outcomes."

Can recommend using functional analogues of many of the above-mentioned ideas as strong practical heuristics.

Not necessarily → **Multi-level consequentialism**
"*Practical decisions* tend to follow established rules or heuristics that generally lead to better outcomes — not case-by-case calculation of expected consequences."

Yes ↓

Single-level consequentialism
An arguably implausible version of pure consequentialism.
"*Case-by-case calculation of expected consequences* is used both as the criterion of rightness and the decision procedure."

Figure 2.1. Minimalist views of wellbeing and value are compatible with a wide variety of normative views, yet are sometimes broadly rejected based on objections to a narrow, implausible combination of views, namely purely 'experientialist welfarist' minimalism combined with '**single-level consequentialism**' (going only straight down in the diagram).

The diagram reflects the structure of this chapter:

- In 2.2, I outline **nonconsequentialist** reasons against absurd acts.
    - I focus on virtue ethics, deontology, social contract theory, care ethics, and skepticism of general moral theories.
- In 2.3, I outline **consequentialist** reasons against absurd acts.
    - I focus briefly on **extra-welfarist** and **extra-experientialist** axiologies, namely on how such views may consider acts of violence or violation to be bad independent of their overall effects on experiential wellbeing.
    - Lastly, I focus on **rule consequentialist** and **multi-level consequentialist** reasons, such as the instrumental reasons for respecting autonomy, cooperation, and nonviolence, which are relevant for all plausible minimalist moral views to the degree that they contain a consequentialist component.

2.2 Nonconsequentialist Reasons Against Absurd Acts

Views of wellbeing alone aren't normative views: they don't in themselves constitute any general principle for us to follow as a 'criterion of rightness' in our moral decision-making. They have normative implications for our actions only when combined with moral views whose criteria of rightness depend on wellbeing. For instance, welfarist consequentialism says that wellbeing outcomes alone determine the rightness of actions, with all other factors — such as intentions, rules, or virtues — being morally relevant only insofar as they affect the wellbeing outcomes.

I assume that all minimalist moral views would give at least some weight to how our actions affect the wellbeing of others. Thus, the consequentialist reasons against absurd acts (outlined in 2.3) can be relevant for all such views. Yet many views may also give additional normative weight to other factors, independent of the consequences of our actions. These views and factors may be seen as separate, nonconsequentialist reasons against absurd acts.[73]
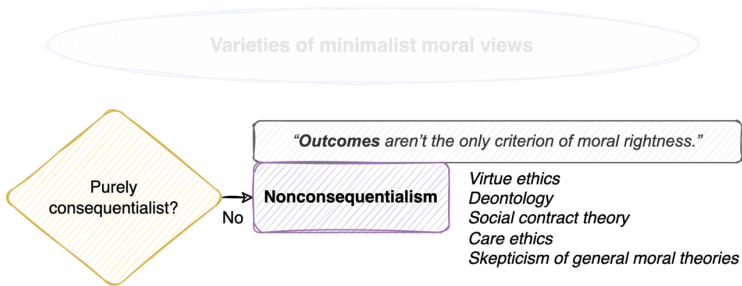
Figure 2.2. The current section covers this branch.

---

[73] For some context regarding how widely endorsed these nonconse-quentialist views are, it is worth noting that among the respondents of the 2020 PhilPapers survey (of 1741 English-speaking philosophers from around the world), only a fifth were exclusively favorable to-ward pure consequentialism (survey2020.philpeople.org/survey/re-sults/4890):

- 37% leaned toward **virtue ethics** (25% exclusively);
- 32% toward **deontology** (20% exclusively);
- 31% toward **consequentialism** (21% exclusively);
- 16% for combined views;
- 6% for alternative views.

The responses among normative ethicists ($n = 358$) indicate similar or even broader support for nonconsequentialist views (survey2020.phil-people.org/survey/results/4890?aos=30):

- 38% leaned toward **virtue ethics** (21% exclusively);
- 41% toward **deontology** (23% exclusively);
- 30% toward **consequentialism** (20% exclusively);
- 22% for combined views;
- 10% for alternative views.

2.2.1 Virtue Ethics

Nothing stands in the way of combining a minimalist view of well-being or value with virtue ethics, whose central focus is the life-long commitment to developing one's moral character. Virtue ethics is not about seeking out any particular actions that yield the best outcomes, but about fostering a virtuous character from which the right actions would naturally follow.[74]

Imagine a moral exemplar who embodies the widely emphasized virtues of courage, kindness, honesty, and integrity. Would they sneak around, opportunistically murdering innocent individuals in the name of reducing suffering? Walk past drowning children? Sabotage healthcare?

They most certainly would not act in such ways. After all, those widely emphasized virtues are highly antithetical to such acts of backstabbing, betrayal, deception, and the like. And even in the unlikely case where one might imagine a consequentialist justification for some seemingly absurd actions, the focus of virtue ethics remains not on any particular actions, but rather on the continuous cultivation of an unfailingly virtuous character, avoiding deviation from the path of highest virtue.[75]

2.2.2 Deontology

A minimalist view may also be combined with deontology, where right action is determined by adherence to a set of moral rules or duties.[76] For instance, deontology can entail a commitment to non-maleficence ("do no harm"), the golden rule ("treat others as you wish to be treated"), or respecting certain inviolable rights that apply universally to all individuals, such as the right to autonomy.

Deontological commitments often directly oppose the kinds of extreme actions that a purely consequentialist analysis might

---

[74] Cf. Hursthouse & Pettigrove, 2023.

[75] As mentioned in 1.3.2.3, one may see virtue as the absence of vice.

[76] Cf. Alexander & Moore, 2021.

otherwise justify as stepping stones to better outcomes. Consider severe rights violations aimed at hastening the termination of lives that are perceived to have negative welfare. Regardless of whether such actions would practically lead to better outcomes, deontology rejects that outcomes are the full picture of what matters morally. Instead, it holds that our actions should primarily align with our duties, which often contradict what may seem justified in the edge cases of purely consequentialist reasoning.[77]

2.2.3 Social Contract Theory

Similarly, one may endorse minimalist versions of social contract theories, which derive moral norms from a hypothetical agreement conceived through rational deliberation (a "social contract"). Social contract theories center around the idea of consensus among rational agents, implying that an action is morally wrong when it violates the norms of this hypothetical consensus.[78]

Imagine a diverse set of people, endorsing a minimalist view of wellbeing, who deliberate on the moral principles governing their society. Would they endorse norms that allow callous acts like murder, passive bystanderism, or attempting to collapse the healthcare system? It seems doubtful that they would endorse the kind of chaotic and unsafe society where such unilateral choices were acceptable.

More likely, they would converge on impartial, predictable norms of justice, trust, and respect for everyone's autonomy, with a focus on cooperatively minimizing severe problems like extreme suffering, violence, and violation. By contrast, the callous acts in

---

[77] Analogous to the case of minimalist virtue ethics: Minimalist deontology would see rule adherence as the absence of rule violation. To fulfill one's duties is to not fail at them, but does not constitute an offsetting moral good.

[78] On contractualism, see Ashford & Mulgan, 2018.
On contractarianism, see Cudd & Eftekhari, 2021.

question seem like textbook examples of breaching the social con-tract.[79]

2.2.4 Care Ethics

Compared to the previous views, care ethics is less abstract and universalistic, and more concrete and contextualistic. A core idea in care ethics is that individuals are fundamentally relational and interdependent beings, with ethical obligations arising from rela-tionships. It is focused on attentive, empathetic, and proactive re-sponsiveness to the needs of others, particularly within relation-ships of care and dependency.[80]

A minimalist version of care ethics would naturally focus on anticipating and addressing the unmet needs of others, without necessarily assuming that their primary concern would be any-thing like minimizing personal suffering. Rather, it would plausi-bly involve attending and responding to others on their own terms, with sensitivity to their own goals and pursuits in life.[81] Thus, minimalist care ethics would likely oppose any acts that fail to

---

[79] While rational agents might converge only on a relatively narrow range of acceptable means for reducing extreme suffering, there's still good reason to assume that they would give strong priority to the under-lying aim of reducing extreme suffering. A contractualism-based argu-ment for the latter claim is found in Vinding, 2020d, sec. 6.7. Similarly, Mayerfeld (1999, p. 115) identifies a contractualist justification for the duty to relieve suffering as follows:

> … reasonable moral rules are those that would be chosen by people made temporarily ignorant of their life circumstances [i.e. behind the 'veil of ignorance']. … People who could not predict the extent of their vulnerability to suffering in real life might seek protection from the worst eventuality by agreeing on a strong requirement to relieve suffering.

[80] Cf. iep.utm.edu/care-ethics.

[81] Given this sensitivity to individuals' own goals and pursuits, it may be natural to combine a minimalist version of care ethics with a mini-malist preference-based view of wellbeing (1.3.2.1).

respond to others in a caring way, such as acts of murder, betrayal, or being insensitive to the subjective perspectives of others.

2.2.5 Skepticism of General Moral Theories

Various philosophers doubt the idea of a universal moral criterion that would apply to all actions in all situations, often highlighting the complexity or context-sensitivity of morality.[82] They may argue that a universal ethical framework will inevitably oversimplify ethics, or that there is no pressing need for such a framework.

A comparison could be made to how physicists use different physical theories in different domains of applicability, reflecting the complexity of physical phenomena. Similarly, one may find some moral theories plausible and applicable in some domains, yet doubt that any single moral theory could fully capture the complexity of all moral phenomena across all domains.

For example, Simon Knutsson combines a minimalist view of wellbeing and value with skepticism of overarching moral theories.[83] The resulting moral view does not generate the alleged absurd implications often falsely attributed to minimalist views of wellbeing or value, as it is not tied to any moral theory that would generate such implications.

· · ·

In sum, while purely consequentialist views deem outcomes the sole criterion of moral rightness, we may also — to the degree that we find it plausible — give independent normative weight to other factors, such as character, duties, agreements, or empathetic responsiveness in our relations with others.[84]

---

[82] Cf. Ridge & McKeever, 2023.

[83] Knutsson, 2023a.

[84] The factors can also be combined in various ways. For instance, one may combine them into "overarching pluralist theories", where factors like different virtues and duties all play an advisory role in all of one's moral decisions. Alternatively, one may combine them into skeptical or

2.3 Consequentialist Reasons Against Absurd Acts

The following reasons against absurd acts relate to varieties of purely consequentialist minimalist views, but also to all other views that give some normative weight to the relevant assumptions, namely:

1. extra-welfarist disvalue,

2. extra-experiential components of wellbeing,

3. rule consequentialism, or

4. multi-level consequentialism.

2.3.1 Axiological Reasons

This section covers views that may consider acts of violence or violation to be bad — in themselves, or for the victim — independent of their overall effects on experiential wellbeing.

2.3.1.1 Extra-Welfarist Axiologies

---

particularist views, where no single aspect of morality is universally applicable, yet many may have their own domains where they best apply.

Figure 2.3. Minimalist versions of ideal consequentialism (sometimes also called ideal utilitarianism) may hold that acts of violence or violation are worth reducing for their own sake.[85]

Even if consequentialist views agree that the rightness of actions is determined solely by the value of outcomes, they diverge in what they define as the morally relevant parts of outcomes. Only purely welfarist views hold that the value of outcomes is based solely on the wellbeing they contain. By contrast, some forms of ideal consequentialism hold that certain acts have intrinsic value or disvalue, independent of their overall effects on wellbeing.[86]

Minimalist versions of ideal consequentialism wouldn't count any acts as intrinsically good or valuable. Yet certain acts would have independent disvalue, thereby decreasing the value of outcomes. For instance, acts like murder or betrayal could in themselves constitute severe bads, and hence count among the very phenomena to be minimized.[87]

2.3.1.2 Extra-Experientialist Welfarist Axiologies
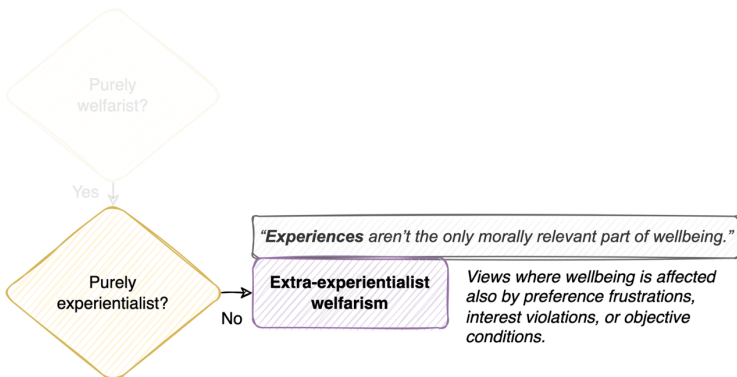
Figure 2.4. One may hold that certain things are bad for us even if they aren't part of our conscious experience.

We have seen how the scope of morally relevant phenomena can be broader than "only consequences" (according to nonconsequentialist views), and how the scope of relevant consequences can be broader than "only effects on wellbeing" (according to extra-welfarist views). Similarly, even if we do assume a purely welfarist consequentialist view, the scope of what we find relevant for wellbeing can be broader than "only conscious experiences".

While experientialist minimalist views (1.3.1) define wellbeing as the degree to which we are free from experiential sources of illbeing (like suffering, disturbance, or a visceral non-acceptance of our current experience), minimalist versions of extra-experientialist views (1.3.2) may additionally hold that we can be severely harmed by factors outside our immediate experience (like unmet preferences, violated interests, or objective conditions).[88]

---

[88] The 2020 PhilPapers survey polled global English-speaking philosophers on wellbeing views. Of the 967 respondents, most seemed to agree that wellbeing can be negatively affected by things outside our experience, as only 10% favored experientialism exclusively (survey2020.philpeople.org/survey/results/5206):

- 53% leaned toward **objective list views** (50% exclusively);
- 19% toward **desire satisfaction/preferentialist views** (15% exclusively);
- 13% toward **hedonism/experientialism** (10% exclusively);
- 5% for combined views;
- 5% for alternative views.

Among the 244 normative ethicists (survey2020.philpeople.org/survey/results/5206?aos=30):

- 64% leaned toward **objective list views** (59% exclusively);
- 15% toward **desire satisfaction/preferentialist views** (12% exclusively);
- 11% toward **hedonism/experientialism** (9% exclusively);
- 6% for combined views;
- 8% for alternative views.

If we combine welfarist consequentialism with these broader views, it follows that we should reduce not only felt harms, but also harms like premature death, failed life projects, and being subjected to violence. This adds another layer of opposition to the absurd alleged implications that people sometimes associate with minimalist views of wellbeing or value.

2.3.2 Rule and Multi-Level Consequentialism

Consequentialist views need not recommend case-by-case calculations of the expected outcomes of every single action. This is often cognitively demanding, time-consuming, or even impossible, and hence practically counterproductive by consequentialism's own lights.

Instead, some versions of consequentialism provide clearer and more practical guidance for action by focusing on general rules or heuristics to follow. These rules or heuristics can capture the wisdom of past experiences, codifying patterns of action that generally lead to better outcomes.

There were weak correlations between consequentialist and experientialist views ($r = 0.26$), between consequentialist and preferentialist views ($r = 0.19$), and an inverse one between consequentialist and objective list views ($r = -0.29$). This indicates that academic philosophers broadly endorse various combinations of normative and wellbeing views, including extra-experientialist versions of consequentialist views. (A minimalist example might be the need-based view of Häyry, 2024, which is about reducing "pain, anguish, and dwarfed autonomy".)

2.3.2.1 Rule Consequentialism
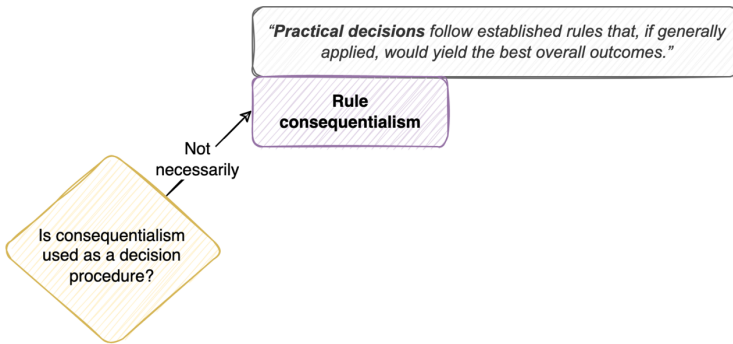
Figure 2.5. Rule consequentialism focuses on the consequences not of individual actions, but of generally applied rules.

Rule consequentialism deems actions right if they follow rules that, when generally applied, yield the best overall outcomes.[89] Unlike deontology and social contract theory, it selects rules based solely on their expected consequences. Yet all three evaluate the moral rightness of individual actions by rule adherence rather than case-by-case consequences.

Minimalist rule consequentialism would strongly oppose acts like unprovoked murder or severe and unprovoked rights violations, as the allowance of such acts is prone to overall increase rather than decrease the amount of problems in the world. At worst, general rules that allowed such acts would risk leading to catastrophic futures. After all, increased conflict and hostility among future actors is a key risk factor for worst-case outcomes — namely, worlds defined by ruthless competition, adversarial dynamics, and escalations that bring out the worst tendencies for hatred, vengeance, and sadism.[90]

Rather, the best general rules to adopt will most likely involve a proactive protection of people's lives and safety, in part because

---

[89] Hooker, 2023.

[90] Baumann, 2019, "Conflict and hostility".

that is arguably the best way to secure and develop our shared ca-
pacity to solve problems in cooperative ways.

2.3.2.2 Multi-Level Consequentialism: Relevant for All Minimalist Moral Views
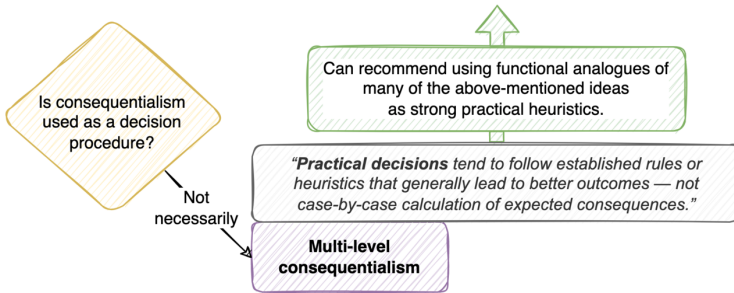
Figure 2.6. The multi-level approach recommends the decision proce-
dures that best help us bring about better outcomes.

Multi-level consequentialism merges act-based and rule-based ap-
proaches, providing a layered approach to consequentialist deci-
sion-making.[91] It ties the rightness of actions to their overall con-
sequences, yet only recommends that we attempt to estimate the
consequences of individual actions in the arguably rare, 'critical'
situations where this is plausibly worth the effort. These could be
the occasional high-stakes situations where our established heuris-
tics deeply conflict, are silent, or might lead to highly suboptimal
outcomes, prompting a switch to the more analytical level of
moral reasoning.

In situations where such detailed analysis is impractical, the
multi-level approach recommends the 'intuitive' decision proce-
dure of following established heuristics that generally lead to bet-
ter outcomes. These heuristics can often be inferred and justified

---

[91] Multi-level consequentialism has more often been discussed under
"two-level consequentialism" or "two-level utilitarianism", as developed
by R. M. Hare.
en.wikipedia.org/wiki/Consequentialism#Two-level_consequentialism;
en.wikipedia.org/wiki/Two-level_utilitarianism.

at the analytical level given our past experiences and knowledge. Yet consequentialists need not reinvent all the moral wheels of society, because a highly sensible meta-heuristic (by consequentialism's own lights) is to assign significant weight to the long-standing recommendations of other ethical views and established norms in typical moral decisions — at least when these recommendations strongly converge to discourage certain types of actions.[92]

If we want to effectively reduce problems in the big picture, the multi-level approach also recommends that we mostly focus on the kinds of positive, constructive goals that best enable us to collectively prevent problems like extreme suffering — a multi-generational, shared endeavor that requires greater levels of coor-

---

[92] Cf. Vinding, 2022j, "A more plausible approach":

> In other words, it seems that utilitarian decision procedures are best approached by assigning a fairly high prior to the judgments of other ethical views and common-sense moral intuitions (in terms of how plausible those judgments are from a utilitarian perspective), at least when these other views and intuitions converge strongly on a given conclusion. And it seems warranted to then be quite cautious and slow to update away from that prior, in part because of our massive uncertainty and our self-deceived minds. This is not to say that one could not end up with significant divergences relative to other widely endorsed moral views, but merely that such strong divergences probably need to be supported by a level of evidence that exceeds a rather high bar.

> Likewise, it seems worth approaching utilitarian decision procedures with a prior that strongly favors actions of high integrity, not least because we should expect our rationalizing minds to be heavily biased toward low integrity — especially when nobody is looking.

> Put briefly, it seems that a more defensible approach to utilitarian decision procedures would be animated by significant humility and would embody a strong inclination toward key virtues of integrity, kindness, honesty, etc., partly due to our strong tendency to excuse and rationalize deficiencies in these regards.

dination and cooperation, and which requires us to avoid and actively prevent absurd acts.[93] And since all the minimalist moral views discussed here give at least some weight to minimizing the badness of outcomes, this is (an added) reason for all people with such views to oppose absurd acts.

---

[93] Vinding, 2022g, chap. 9, "Identifying Plausible Proxies"; forthcoming, sec. 6.3, "Focusing on Positive and Constructive Goals".

Part II

Minimalist Views in Theory and Practice

Chapter 3

Minimalist Axiologies and Positive Lives

Minimalist views of value (axiologies) are evaluative views that define betterness solely in terms of the absence or reduction of independent bads, such as suffering. This chapter looks at minimalist axiologies that are impartial and welfarist (i.e. concerned with the welfare of all sentient beings), with a focus on their theoretical and practical implications. For example, these views reject the 'Very Repugnant Conclusion' implied by many offsetting ('good minus bad') views in population ethics.

Minimalist views are arguably neglected in population ethics due to their apparent implication that no life could be axiologically positive. After all, minimalist views reject the concept of independent goods. Yet these views are perfectly compatible with the notion of relational goods, and can thereby endorse relationally positive lives that make a positive difference for other beings.

This notion of relationally positive value is entirely excluded by the standard, restrictive assumption of treating lives as isolated

value-containers. However, this assumption of 'all else being equal' is practically always false, enabling the possibility of highly positive lives according to minimalist views.

Minimalist views become more intuitive when we adopt a relational view of the overall value of individual lives, that is, when we don't track only the causally isolated "contents" of these lives, but also their (often far more significant) causal roles.

3.1 What Is Axiology?

Axiology is the philosophical study of value.[94] Here, I will focus on questions related to independent value, relational value, and resolving conflicts between values.

3.1.1 Independent Versus Relational Value

Axiology is centrally concerned with the question of what things, if any, have independent value, also known as intrinsic value.

'Axiologies' in the plural refer to specific views on this axiological question. Once we assume a specific axiology — that is, a view that ascribes independent value to certain entities or states — we may then understand the value of all other things as extrinsic, instrumental, or relational in terms of their effects on these entities or states.

This distinction between independent and relational value applies at the level of our axiological theory. The distinction can blur at the level of our everyday perception, and this blurring is often practically adaptive.

For instance, we may both formally deny that something has *independent* value, yet also correctly perceive that it does have *value*, without constantly unpacking what this value relationally depends on. Our decision-making tends to be more efficient when we perceive the various objects of our goals as simply having

---

[94] Schroeder, 2018.

value, even when these goals are only indirectly related to what we ultimately value.

Thus, we may do well in practice by treating many widely-held values — such as autonomy, health, or friendship — as independently valuable things to safeguard and promote, at least until they run into conflicts with each other. When they conflict, we have reason to clarify what the underlying or ultimate source of their value might be.[95]

3.1.2 Resolving Conflicts Between Values

To figure out what things have independent value, we commonly devise thought experiments where only a single thing is meant to be changing, all else being equal. We then ask our supposedly value-tracking intuitions whether it seems true that this isolated change is accompanied by a change in value.

Based on such thought experiments of isolated value, one might come to endorse at least two, independent standards of value: "The more positive pleasure (or bliss), the better", and "The less agony, the better".

At first glance, these two standards of value may seem to be perfectly compatible with each other, given that the more blissful a mind-moment is, the less agonized it is. Yet dilemmas arise if we want to simultaneously follow both standards more widely, as they are not always perfectly anticorrelated. That is, we often cannot *both* "maximize bliss" *and* "minimize agony" at the same time, because even as these two guiding principles may seem to be polar opposites, they do not always constitute a coherent twin-principle like "Head North, Avoid South".

The field of population ethics has highlighted ways in which these principles come apart, pulling our intuitions into mutually

---

[95] Cf. the topic of value commensurability as discussed in 1.1.2 and in Schroeder, 2018. (More on everyday perception and practical heuristics in 2.3.2.2; 3.5.2; 6.2.)

incompatible directions. And it has highlighted the lack of consensus about how to compare the intrinsic value of positive pleasure against the intrinsic disvalue of agony.[96]

To resolve conflicts between the seemingly intrinsic dual values of positive pleasure versus agony, one option is to establish acceptable tradeoff ratios (or "priority weights") between them, so as to clarify how much a change in one weighs against a change in another.

Another approach is to reject the assumption that there are any truly independent goods to begin with. This is the approach of minimalist axiologies, where the value of purported goods like bliss is weighed in terms of how well they reduce bads like agony. (Note that according to some views, such as views in the Epicurean tradition, bliss is understood as the complete absence of any pain or unpleasantness, and hence "maximizing bliss" and "minimizing unpleasantness" are indeed equivalent twin-principles on this conception of bliss.[97])

3.2 What Are Minimalist Axiologies?

3.2.1 The Less This, the Better

Minimalist axiologies essentially say: "The less this, the better". In other words, their fundamental standard of value is about the *avoidance* of something, and not about the maximization of something else.

To list a few examples, minimalist axiologies may be formulated in terms of avoiding…

---

[96] Many of the relevant thought experiments in population ethics will be visualized shortly in 3.3 and Chapter 4.

[97] 1.3.1.2; cf. Sherman, 2017; Knutsson, 2019, 2022b.

- cravings (tranquilism[98]; certain Buddhist axiologies[99]);

- disturbances (Epicurean minimalism[100]);

- pain or suffering (Schopenhauer[101]; Richard Ryder[102]);

- frustrated preferences (antifrustrationism[103]); or

- unmet needs (Häyry[104]; some interpretations of care ethics[105]).

This chapter looks at minimalist axiologies that are impartial[106] and welfarist[107]: focused on the welfare of all sentient beings.[108]

---

[98] 1.3.1.4; Gloor, 2017.

[99] Anonymous, 2015, sec. 2.2; Breyer, 2015.

[100] 1.3.1.2; cf. Sherman, 2017; Knutsson, 2019, 2022b.

[101] Schopenhauer, 1819, "the negative nature of all satisfaction"; 1851, "negative in its character"; Fox, 2022, "pleasures of distraction".

[102] Ryder, 2001, p. 27.

[103] 1.3.2.1; Fehige, 1998. See en.wikipedia.org/wiki/Antifrustrationism.

[104] Häyry, 2024.

[105] Cf. iep.utm.edu/care-ethics. Specifically, a minimalist interpretation of care ethics could say that our main responsibility is to ensure that there are fewer unmet needs, and not to create additional needs at the cost of neglecting existing or expected needs.

[106] On moral impartiality, see Jollimore, 2023.

[107] en.wikipedia.org/wiki/Welfarism.

[108] Thus, the chapter is not meant to cover axiologies that might be technically minimalist, yet which are partial or focused on non-welfarist avoidance goals (like minimizing human intervention in nature, or avoiding the loss of unique information).

3.2.2 Relational Value in Light of Impartial Avoidance Goals

In tradeoffs between multiple, seemingly independent values, minimalist axiologies avoid the issue of having to establish independent priority weights for different goods and bads so as to resolve their mutual conflict. (By contrast, offsetting axiologies face this issue in tradeoffs such as creating bliss for many at the cost of agony for others.)

Instead of using different standards of value for goods and bads, minimalist axiologies construe 'positive value' in a purely relational way, with regard to an overall avoidance goal for all beings. This transforms the apparent conflict between goods and bads into an empirical question about the degree to which the goods can reduce the bads, and thus still be genuinely valuable in that way.

When we look at only one kind of change in isolation, it may seem intuitive that bliss is independently good and agony independently bad. Yet we may also feel internally conflicted about tradeoffs where value and disvalue need to be compared with each other, so that we could say whether some tradeoff between them is "net positive" or not. Put differently, we may have both promotion intuitions and avoidance intuitions that seem to lack a common language.

To solve these dilemmas, minimalist axiologies would respect promotion intuitions to the degree that they are conducive to the overall avoidance goal, yet reject the creation of more (isolated) value for some at the cost of disvalue for others.[109]

For example, suffering-focused minimalism would respect the promotion of wellbeing *in the place of* suffering, but not of wellbeing *at the cost of* suffering, all else equal.[110]

---

[109] More on the assumption of causally isolated value in 3.4.

[110] Cf. Vinding, 2020d, chap. 3.

Thus, minimalist axiologies sidestep the problem of having to find acceptable "tradeoff ratios" between independent goods versus bads, replacing it with the relational question of how the objects of our promotion intuitions could help with the overall avoidance goal.

3.2.3 Contents Versus Roles

Minimalist axiologies may appear to imply that "No life could be positive or worth living". And such a conclusion might seem implausible to our intuitions, which might say that "Surely lives can be positive or worth living".

Yet minimalist axiologies merely imply that individual lives cannot have positive value when understood as "isolated value-containers", which they never are in the real world.[111] And given that the "no life could be worth living" conclusion only follows when we adopt this highly unrealistic isolated view of individual lives, our intuitive objection to this conclusion need not stem from the sentiment that "Surely lives can have at least some *isolated* positive value". Instead, we might reject the conclusion because we reject, or fail to accurately imagine, its unrealistic premise of treating lives as isolated value-containers.

According to minimalist views, positive value is not something that we "have", "contain", or "accumulate" in isolation, but rather something that we "do" for a wider benefit. This relational view of positive value may conflict with the Western cultural tendency[112] to see individuals as "independent, self-contained, autonomous entities"[113] and to ascribe positive value to the collection of particular experiences for their own sake. Thus, minimalist views may be neglected by the heavily Western-influenced field

---

[111] More in 3.4.

[112] Henrich, Heine, & Norenzayan, 2010b, "Independent and interdependent self-concepts".

[113] Markus & Kitayama, 1991; Remmelt, 2021, "We view individuals as independent".

of population ethics, where we often draw unrealistically tight boxes around individual lives in our quest to isolate that which makes a life valuable or worth living.

To better explore our intuitions about supposedly isolated positive value without the confounding influence of our positive roles-tracking intuitions, it may be helpful to explicitly imagine that we are offered the chance to create a causally isolated black box whose existence has no effects beyond itself. How positive could the box be? Minimalist views say that it cannot be positive at all, regardless of what it contains.

With all that said, the next section will consider the implications of minimalist versus offsetting views when we do play by the rules of population ethics and draw boxes around the contents of individual lives. Overall, many counterintuitive conclusions in population ethics may be attributed to the view of positive value as an independent and independently aggregable phenomenon, or a "plus-point" that can be summed up or stacked in isolation from the positive roles of the individual lives or experiences that contain it.

3.3 How Do Minimalist Views Help Us Make Sense of Population Ethics?

Population ethics is "the philosophical study of the ethical problems arising when our actions affect *who* is born and *how many* people are born in the future".[114] A subfield of population ethics is *population axiology*, which is about figuring out what makes one state of affairs better than another. Some have argued that this is a tricky question to answer without running into counterintuitive conclusions, at least if we make the assumption of *independently* positive lives.[115] Minimalist axiologies do not make this

---

[114] en.wikipedia.org/wiki/Population_ethics.

[115] Cf. Arrhenius, 2000a. For a counterpoint to the common narrative that population ethics is particularly difficult, see Knutsson, 2024.

assumption, and hence they avoid the conclusions that are pictured in the three diagrams in the next three subsections.[116]

Before looking at the diagrams, let us already note a way in which people might implicitly disagree about how to interpret them. Namely, some of the diagrams contain a horizontal line that indicates a "zero level" of "neutral welfare", which may be interpreted in different ways. For example, when diagrams illustrating the (Very) Repugnant Conclusion contain lives that are "barely worth living", some may think that these lives involve "slightly more happiness than suffering", while others may think, as Derek Parfit originally did, that they "never suffer".[117]

A different interpretation of the horizontal line is used in antifrustrationism by Christoph Fehige, where welfare is defined as the avoidance of preference dissatisfaction (or 'frustration'). When Fehige's own diagrams contain the horizontal line, it just means the point above which the person has "a weak preference for leading her life rather than no life".[118] On Fehige's view, the lives with "very high welfare" are much better off than the lives "barely worth living" that still contain a lot of frustration.

Yet if we assume that the lives above the horizontal line have *all* their preferences satisfied, "never suffer", and have no bads in

---

[116] The conclusions are named "paradoxical" or "repugnant" after the intuitions of people who find them troubling. Generally, people differ a lot in which intuitions they are willing to "give up" in population ethics.

[117] As noted by Anthony DiGiovanni (2021a) on Parfit's original formulation of the Repugnant Conclusion (RC):

> [Parfit] explicitly says the beings in this world "never suffer." Many suffering-focused axiologies would accept the RC under this formulation—see e.g. Wolf (1997)—which is arguably a plausible conclusion rather than a "repugnant" bullet to bite. However, in many common formulations of the RC, the distinguishing feature of these beings is that their lives are just barely worth living according to axiologies other than strongly suffering-focused ones, hence they may contain a lot of suffering as long as they also contain slightly more happiness.

[118] Fehige, 1998, p. 534.

their lives whatsoever, then minimalist axiologies would find no problem in the Mere-Addition Paradox or the Repugnant Conclusion. Even so, they would still not *strictly* prefer larger populations, finding all populations of such problem-free lives rather equally perfect (in causal isolation).

However, it seems unusual to imagine that the lives "barely worth living" would be subjectively completely untroubled or "never suffer". Thus, we will here assume that the lives just above the line are not completely untroubled, as seems common in population ethics.[119]

3.3.1 The Mere-Addition Paradox

Derek Parfit's Mere-Addition Paradox (Figure 3.1) is based on a comparison of four populations. Each bar represents a distinct group of beings. The bar's width indicates their numbers, and the height their level of welfare. We assume that every being in this diagram has "a life worth living". (The populations in *A+* and *B−* consist of two isolated groups; the population in *B* is simply the two groups of *B−* combined into one.)
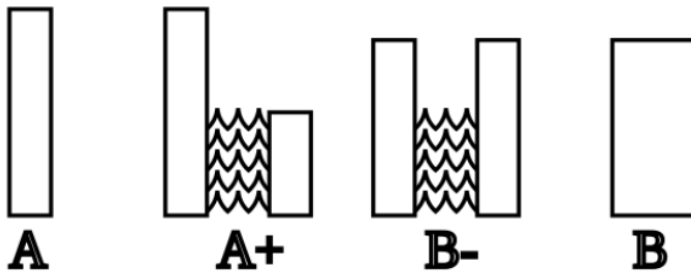
Figure 3.1. The Mere-Addition Paradox.[120]

---

[119] See, for instance, Knutsson, 2024, sec. 4.4.

[120] From en.wikipedia.org/wiki/Mere_addition_paradox.

The paradox results from the following evaluations that to-gether contradict some people's intuitive preference for the high-average population of *A* over the lower-average population of *B*:

1. Intuitively, "*A+* is no worse than *A,*" since *A+* simply con-tains more lives, all worth living.

2. Next, "*B−* is better than *A+*," since *B−* has *both* greater total welfare *and* greater average welfare.

3. Finally, "*B−* is equal to *B,*" since *B* is simply the same groups, only combined.

4. Now, "*B* is better than *A,*" based on steps 1–3.

This paradox is a problem for those who strongly feel that "*A* is better than *B*", yet who are also sympathetic to total utili-tarianism. One reason to avoid average utilitarianism or averagism is that it implies "sadistic conclusions", in which average welfare is increased by the addition of hellish lives.[121] Yet if we assume that the lives in *B* contain more subjective problems than do the lives in *A*, then minimalist axiologies would prefer *A* over *B* without averagism.

Essentially, Fehige's solution is to assume that the welfare of a life depends entirely on its level of preference dissatisfaction or 'frustration'. On this view, a population of satisfied beings cannot, other things being equal, be improved by the "mere addition" of new, less satisfied beings. This is because the frustration of those new beings is an additional subjective problem, as compared to the non-problematic non-existence of their imaginary counterparts in the smaller population.[122]

---

[121] Cf. Parfit, 1984, p. 422, "Hell Three"; Arrhenius, 2000a, p. 251, "The Sadistic Conclusion".

[122] Fehige, 1998. Fehige's use of the term 'preference frustration' is much broader than the everyday feeling that we call frustration. After all, basically all lives in the real world have at least some of their preferences frustrated, even if some may be free from the feelings of frustration.

While this may be a theoretically tidy solution to the Mere-Addition Paradox, critics have objected that it depends on a theory of welfare that they find to be counterintuitive, incomplete, or unconvincing.

However, we would be wise to abstain from hastily dismissing minimalist views as being counterintuitive, because there are plenty of ways to interpret them in more intuitive ways without losing their theoretical benefits. A lot of their perceived incompleteness might result from the thought experiments themselves, which imply that we are not actually supposed to imagine the kind of lives that are familiar to us, but only the hypothetically isolated kind of lives that have absolutely no positive roles beyond themselves.

3.3.2 The Repugnant Conclusion

By continuing the logic of "mere addition", we arrive at the 'Repugnant Conclusion' (Figure 3.2):

> In Derek Parfit's original formulation[,] the Repugnant Conclusion is characterized as follows: "For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better even though its members have lives that are barely worth living" (Parfit 1984). … The Repugnant Conclusion is a problem for all moral theories which hold that welfare at least matters when all other things are equal.[123]

---

[123] Arrhenius, Ryberg, & Tännsjö, 2014.
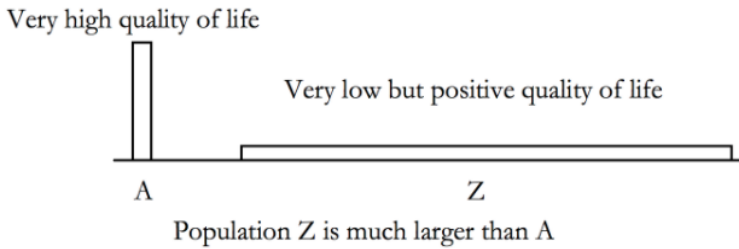
Figure 3.2. The Repugnant Conclusion.[124]

Minimalist axiologies avoid the Repugnant Conclusion, as they deny that the lives "barely worth living" would constitute a vast heap of independent "plus-points" in the first place. For example, Fehige would assume that the lives with "a very high quality of life" would be quite free from problems, which is better, all else equal, than a much larger set of lives that still have a lot of their preferences unsatisfied.

As noted by another commenter on Fehige's paper:

> Among its virtues, [antifrustrationism] rescues total utilitarianism from the repugnant conclusion. If utility is measured by the principle of harm avoidance instead of aggregated preference satisfaction, utilitarianism does not, as the accusation often goes, entail that it is better the more (acceptably) happy lives there are[, other things being equal].[125]

Fehige's theory of welfare is seemingly dismissed by some philosophers on the grounds that it would, counterintuitively, deny the possibility of lives worth living:

> However, a theory about welfare that denies the possibility of lives worth living is quite counter-intuitive [Ryberg, 1996]. It implies, for example, that a life of one year with complete preference satisfaction has the same welfare as

---

[124] From Arrhenius, Ryberg, & Tännsjö, 2014.

[125] Karlsen, 2013, p. 160.

> a completely fulfilled life of a hundred years, and has higher welfare than a life of a hundred years with all preferences but one satisfied. Moreover, the last life is not worth living (Arrhenius 2000b).[126]

Yet this objection seems to imply that a life could be worth living *only* for its own sake, namely for some kind of satisfaction that it independently "contains", and to deny that a life could be worth living for its positive roles. Again, we need to properly account for the fact that Fehige's model is only comparing lives in causal isolation.

As soon as we step outside of the hypothetical, isolated case where "other things are equal" and start comparing these lives in our actual, interpersonal world, we may well see how, even on minimalist terms, the subjectively perfect one-year life could be much less valuable (overall, for all beings) than would be the subjectively near-perfect century. After all, many of our preferences and preferred actions have significant implications for the welfare of others.

However, it is not necessarily counterintuitive to prefer the perfect year, or even non-existence, over the imperfect century in the hypothetical case of complete causal isolation, where we can make no positive difference in any way. Regardless of how we felt during the year, or during the century, others would live as if we never had. Thus, we may question the overall worth of extending our life solely for our own sake in the experience machine, or in an equally solipsistic preference satisfaction machine, provided that it solves no problem beyond ourselves.

3.3.3 The Very Repugnant Conclusion

The Repugnant Conclusion was termed repugnant due to the intuition that a legion of lives "barely worth living" cannot be better than a smaller population of lives that each have a very high welfare. Some say that in this case, the intuition is wrong and that we

---

[126] Quote from Arrhenius, Ryberg, & Tännsjö, 2014, sec. 2.4.

should "bite the bullet" and follow the utilitarian math of additive aggregationism.[127] However, presumably fewer people would accept the 'Very Repugnant Conclusion' (VRC), in which the better world, according to many offsetting views, contains a lot of subjectively hellish lives, supposedly "compensated for" by a vast number of lives that are barely worth living (Figure 3.3):

> There seems to be more trouble ahead for total [offsetting] utilitarians. Once they assign some positive value, however small, to the creation of each person who has a weak preference for leading her life rather than no life, then how can they stop short of saying that some large number of such lives can compensate for the creation of lots of dreadful lives, lives in pain and torture that nobody would want to live?[128]
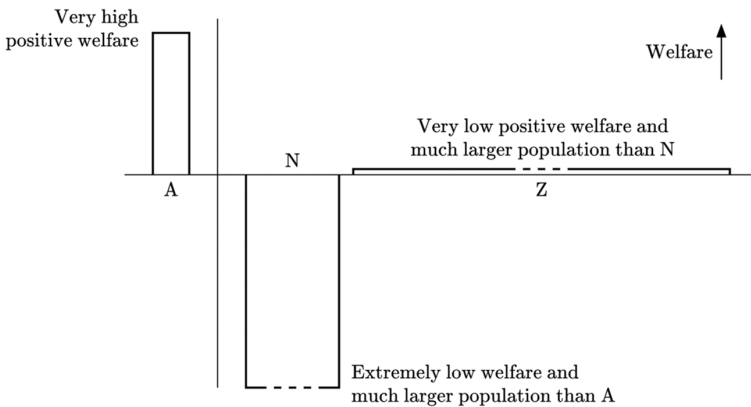
Figure 3.3. The Very Repugnant Conclusion (VRC).[129]

In other words:

---

[127] Additive aggregation entails that the total value of a group or state of affairs is simply the sum of the individual value of each of its members.

[128] Fehige, 1998, pp. 534–535.

[129] From Tännsjö, 2020.

> Let W1 be a world filled with very happy people leading meaningful lives [**A**]. Then, according to total [offsetting] utilitarianism, there is a world W2 which is better than W1, where there is a population of [purely] suffering people [**N**] much larger than the total population of W1, and everyone else has lives barely worth living [**Z**] - but the population is very huge.[130]

One way to avoid the VRC is to follow Fehige's suggestion and interpret utility as "a measure of avoided preference frustration". On this view, utilitarianism "asks us to minimize the amount of preference frustration", which leads us to prefer *W1* over *W2*.[131] As noted by Fehige, "Maximizers of preference satisfaction should instead call themselves minimizers of preference frustration."[132]

Every minimalist axiology would prefer *W1* over *W2* due to being structurally similar to Fehige's view — that is, none of them would say that the supposed "plus-points" of *W2* could somehow independently "counterbalance" the agony of the others, regardless of the number of the lives "barely worth living".

In contrast, the VRC is a problem for many offsetting axiologies besides purely hedonistic ones:

> Consider an axiology that maintains that any magnitude of suffering can be morally outweighed by a sufficiently great magnitude of preference satisfaction, virtue, novelty, beauty, knowledge, honor, justice, purity, etc., or some combination thereof. It is not apparent that substituting any of these values for happiness in the VRC makes

---

[130] From Armstrong, 2019. Formal discussions of the VRC are sometimes traced back to Arrhenius, 2003. However, the VRC was discussed already in Fehige, 1998.

[131] Fehige, 1998, pp. 535–536.

[132] Fehige, 1998, p. 518.

> it any more palatable[.][133]

Minimalist views also avoid what are arguably even stronger objections against offsetting total views, such as the theoretical choices of 'Creating hell to please the blissful' (Figure 4.5) and 'Intense bliss with hellish cessation' (Figure 5.2). More on the comparative implications between minimalist and offsetting views in the next chapters.

3.3.4 Solving Problems: A Way to Make Sense of Population Ethics?

In general, a way to avoid the VRC (and the two other conclusions above) is to hold that ethics is about solving and preventing problems, and not about creating new, unneeded goods elsewhere for their own sake. On this view, any choice between two populations (all else equal) is based on considering which population contains the overall greater amount of problematic states, such as extreme suffering.

This problem-focused view rejects the metaphor that ethical problems could be "counterbalanced" instead of prevented:

> [Only] the existence of such problematic states imply genuine victims, while failures to create supposed positive goods (whose absence leaves nobody troubled) do not imply any real victims — such "failures" are mere victimless "crimes". … According to this view, we cannot meaningfully "cancel out" or "undo" a problematic state found somewhere by creating some other state elsewhere.[134]

Generally, the metaphor of ethical counterbalancing may stem from our common tendency to think in terms of polar opposites. When we speak of a 'negative' state, we may naturally assume

---

[133] DiGiovanni, 2021a, sec. 1.1.1.

[134] Vinding, 2020c, sec. 2.4.

that it could be counterbalanced by a 'positive' state of equal magnitude.[135] Yet the mere observation of a negative state does not imply the possibility of a corresponding positive or 'anti-negative' state.[136] The opposite of a problematic state may, instead, be just an unproblematic state, with no equivalent 'anti-problematic' state to counterbalance it.[137]

3.4 What Are We Comparing When We Make the Assumption of "All Else Being Equal"?

3.4.1 Isolated Value-Containers

The *ceteris paribus* assumption is often translated into English as something like "all else equal", "all else unchanged", or "other things held constant".[138] That is, we exclude any changes other than those explicitly mentioned. When we make this assumption in population ethics, the idea is to compare any two hypothetical populations only with respect to their explicit differences, such as the level and distribution of welfare among those populations, with no other factors influencing our judgment.

Yet we need to be mindful of the potential pitfalls when we compare populations in this way. For instance, it is much easier said than done to completely rule out the influence of all factors other than those explicitly mentioned. We may *think* we have done it upon reading the words "all else equal", yet we may in fact need to spend some time and imaginative effort to actually prevent such supposedly external factors from influencing our judgment.

---

[135] Vinding, 2020d, pp. 155–156; Knutsson, 2021b, sec. 3; Leighton, 2024.

[136] Vinding, 2022e.

[137] Cf. Figure A1.1.

[138] en.wikipedia.org/wiki/Ceteris_paribus.

This pitfall is exacerbated by the fact that the "all else equal" assumption is often left highly implicit, with little attention paid to how radically it changes what we are talking about. Quite often, we only see it as a parenthetical (as if ignorable) remark, with no instruction for how to best account for it. Or we may not see it at all, as in situations where it is an unvoiced background assumption, in which we are trusted to have the contextual awareness to take it properly into account even with no explicit reminder to do so.

How we account for the *ceteris paribus* assumption can influence the distance that we see between theory and practice, which in turn can influence our potentially action-guiding views, such as our views on what kinds of lives can be positive or worth living (and what for).

To illustrate, let us imagine a scenario where the *ceteris paribus* assumption would actually be true: namely, we are comparing only "isolated value-containers" or "isolated Matrix-lives" that never interact with each other in any way.[139] If this sounds radical, then we may not always realize how radical the assumption in fact is. After all, we may intuitively assume that "positive lives" would also play positive social roles and make a difference beyond themselves. Yet these factors are supposed to be ruled out when we are comparing lives *ceteris paribus*, solely for their own sake.

In effect, when we apply the *ceteris paribus* assumption to the value of individual lives, we are restricted to a kind of "isolated view" of lives worth living, as if that which makes a life worth living would necessarily have to be something that it contains, rather than its roles and relations beyond itself.

3.4.2 Counterintuitive Boundaries

Our practical intuitions about the overall value of lives — such as of all the lives "barely worth living" in the (Very) Repugnant Conclusion — may implicitly be tracking not only the "contents" of

---

[139] Not even by acausal "influence".

these lives (i.e. their own level of welfare), but also their overall effects on the welfare of others.

And in practice, it may indeed seem like a repugnantly bad idea to trade away a high-welfare population for a legion of lives "barely worth living", as the latter might seem to not have enough wellbeing as a resource to adequately take care of each other in the long term. (A practical intuition in the opposite direction is also possible, namely that a larger population could create more goods, insights, and resources that everyone could benefit from, and thus have a brighter future in the long run.)

Yet to give any weight to such instrumental effects, even implicitly, would already conflate our evaluation of those hypothetical lives solely for their own sake. After all, we are supposed to compare only the level and distribution of welfare as shown in the population ethics diagrams, and to ignore our practical intuitions about how the lives or populations might evolve or unfold in different ways if interpreted as the starting point of a story in the real world.

Our practical intuitions are adapted for an interpersonal world with a time dimension: two features of life that are difficult for us to put aside when entering thought experiments about the value of other beings. Thus, we may need to explicitly remind ourselves that the population ethics diagrams are, in effect, already depicting the complete outcome, with no relevant interactions or time-evolutions left outside the box.

3.5 What Do These Views Imply in Practice?

3.5.1 Naive Versus Sophisticated Minimalism

Even if minimalist views avoid many of the conclusions that have been called tricky problems for other views in population ethics,

one might still worry that minimalist views could have absurd implications in practice. Yet regardless of what specific implications one has in mind, it is worth noting that many of them might stem not only from the "isolated view" that ignores the positive roles of individual lives, but also from the hasty logic of a "naive" form of consequentialism, which further ignores the positive roles of various widely established social norms, such as those of respecting autonomy, cooperation, and nonviolence.[140]

A naive consequentialism is not based on a nuanced view of expected value thinking,[141] and can instead fall victim to a kind of "narrative misconception" of consequentialism, in which the view would support "any means necessary" to bring about its axiologically ideal "end state".

One could argue that the idea of a 'utilitronium shockwave' (i.e. turning all accessible matter into pure bliss) amounts to such a misconception about the practical implications of classical utilitarianism.[142] In the case of minimalist axiologies, this misconception looks like the claim that we must, at any cost, "seek a future where problems are eventually reduced to zero", which is very different from minimizing the amount of problems over *all* time in expectation.

After all, only the second kind of thinking — namely, expected value thinking without any fixed destination — is sensitive to risks of making things worse. By contrast, the first, misconceived view is more like fixating on a particular story of what we must eventually achieve at some particular time in the future.

And instead of being sensitive to risks of making things worse, the story might include a point, as many stories do, at which the protagonists must engage in an "all in" gamble to ensure that they bring about an ideal world "in the end". In other words,

---

[140] More in Chapter 2 and in 6.3.

[141] Cf. Todd, 2021; probablygood.org/core-concepts/expected-value.

[142] For the utilitronium shockwave thought experiment by David Pearce, see hedweb.com/social-media/pre2014.html.

the narrative misconception of consequentialism might hold that "what justifies the means" is the endpoint, or the possibility of reaching it, rather than the overall minimization of problems with no fixed destination.

More concretely, a naive version of minimalism might lead us to ignore the positive norms of everyday morality as soon as there would (apparently) be even the slightest chance of bringing about its hypothetically ideal "end state", such as an empty world, even if doing so would violate the preferences of others or risk multiplying the amount of problems in the future by many orders of magnitude.

By contrast, a nuanced or "sophisticated" version of minimalism would be concerned with the "total outcome" — which spans all of time — and be highly sensitive to the risk of making things worse overall. For instance, any aggressively violent strategy for "preventing problems" would very likely backfire in various ways, such as by undermining one's credibility as a potential ally for large-scale cooperation, ruining the reputation of one's (supposedly altruistic) cause, and eroding the positive norm of respecting individual autonomy.[143]

Given that the backfire risks depend on complex interactions that happen over considerable spans of time, we are likely to pay them insufficient attention if our thinking of real-world interventions is as simplistic as the boxes that collapse hypothetical populations into two-dimensional images. Of course, a nuanced, practical minimalism would not be like thinking in terms of boxes, and would instead take the relational factors and empirical uncertainties into account.[144]

---

[143] Cf. 6.3.1; Vinding, 2020e.

[144] More on the "narrative misconception" of consequentialism in 5.3.

3.5.2 Compatibility with Everyday Intuitions

What, then, do these views imply in practice, assuming a sophisticated minimalism over *all* time? The second half of Magnus Vinding's *Suffering-Focused Ethics* is an accessible and extensive treatment of basically the same question, particularly for views that prioritize minimizing extreme suffering.[145] That is, the practical implications of minimalist views are probably in large part the same as those of many suffering-focused views.

Yet minimalist views differ from at least some suffering-focused views in one respect, which is that minimalist views work completely without the concept of independent positive value, placing full emphasis on relational positive value.

For that reason, minimalist views may appear as if they were somehow uniquely opposed to many things that we might intuitively cherish as being intrinsically valuable — as if none of our intuitively positive pursuits would have any positive worth or weight to justify their inevitable costs.

Yet minimalist views need not imply anything radical about the *quantity* of positive value that we intuitively attribute to many things at the level of our everyday perception.

After all, the kinds of things that we may deem "intrinsically valuable" at an intuitive level are often precisely the kinds of things that rarely need any *extrinsic justification* in everyday life, such as sound physical and mental health, close relationships, and intellectual curiosity.

If required, we often *could* "unpack" the value of these things in terms of their indirect, long-term effects, namely their usefulness for preventing more problems than they cause. But when our intuitively positive pursuits have many beneficial effects across a variety of contexts, we are often practically wise to avoid spending the unnecessary effort to separately "unpack" their value in relational terms.

---

[145] Vinding, 2020d, pp. 141–277.

Additionally, the more we unpack and reflect on the relational benefits of our intuitively positive pursuits, the more we may realize the full magnitude of their positive value, even according to minimalist views. After all, if our perception and attribution of positive value is focused on our positive feelings in the immediate moment, we may actually underestimate the overall usefulness of things such as maintaining good health and relationships, learning new skills, and coming up with new insights. Namely, it is of course desirable when such things help us feel better, yet perhaps the bulk of their value is not how they affect our own feelings in the moment, but what roles they play for all beings.

In other words, while minimalist views may not assign positive value to any particular experiences solely for their own sake, they can still value all the physical, emotional, intellectual, and social work, as well as any other activities, that best enable us collectively to alleviate problems for all beings.

Overall, if the goal is to minimize problems, we are faced with the dauntingly complex meta-problem of identifying interventions that can reasonably be expected to prevent more problems than they cause. And this meta-problem will require us to combine a vast amount of knowledge and *supportive* values. That is, minimalist views do not imply that we hyper-specialize in this meta-problem in a way that would dismiss all seemingly intrinsic values as superfluous. Rather, they imply that we adhere to a diverse range of these values so as to advance a mature and comprehensive approach to alleviating problems.[146]

3.5.3 Preventing Instead of Counterbalancing Hell

Of course, it would be a 'suspicious convergence'[147] if all the things that we may perceive as being intrinsically valuable would

---

[146] For more on what minimalist views might support in practice, see 2.3.2.2; 5.3; and Chapter 6.

[147] Cf. Lewis, 2016.

also be relationally aligned with the impartial minimization of problems. Yet the everyday implications of minimalist views need not be very different from those of other consequentialist views, as all of them imply a personal ideal of living an effective life in alignment with some overall optimization goal — which, in turn, may recommend a broadly shared set of habits and heuristics for everyday living.

Where the views differ the most may be in their large-scale implications. For instance, instead of primarily ensuring that we expand out into space, minimalist views would imply that we prioritize steering the future away from worst-case scenarios. After all, many scenarios of space colonization may, depending on their guiding values, vastly increase the amount of suffering over all time (in expectation).[148]

3.5.4 Self-Contained Versus Relational Flourishing

When psychologists speak of flourishing, it can have many meanings. As a value-laden concept, it is often bundled together with things like "optimal growth and functioning", "social contribution", or "having a purpose in life".[149]

Before we load the concept of flourishing with independent positive value, as is seemingly done by the authors of works such as Utilitarianism.net[150], *The Precipice*[151], and *What We Owe the Future*,[152] it is worth carefully considering whether this value is best seen as independent or relational in the axiological sense.

Minimalist views would not see positive flourishing as any kind of "self-contained" phenomenon of isolated value for one's

---

[148] Tomasik, 2013d.

[149] en.wikipedia.org/wiki/Flourishing.

[150] Chappell, Meissner, & MacAskill, 2021, "flourish".

[151] Ord, 2020, "flourish".

[152] MacAskill, 2022, "flourish".

own sake. Yet minimalist views are perfectly compatible with a relational notion of positive flourishing as personal alignment with something beyond ourselves. For instance, minimalist flourishing could mean that we are skillfully serving the unmet needs of all sentient beings, aligning our wellbeing with the overall prevention of illbeing.

In practice, a minimalist understanding of "optimal growth and functioning" would probably entail a combination of strategic self-investment and healthful living (similar to other impartial welfarist views). After all, we first need to patiently grow our strengths, skills, and relationships before we can sustainably and effectively apply ourselves to help others. And because life can be long, it makes sense to keep growing these capacities, meeting our needs in harmony with the needs of others, and to seek and find the best ways for us to play positive roles for all sentient beings.[153]

3.6 Without the Concept of Intrinsic Positive Value, How Can Life Be Worth Living? A More Complete View

In standard, theoretical population ethics, what we see are only the isolated "welfare bars" of what all the individual lives independently "contain". Yet in practice, we also have hidden "relational roles bars" of what our lives "do" beyond ourselves.

On *any* impartial and welfarist view, our own "aggregate welfare" is often a much *smaller* part of our life's overall value than is what we do for the welfare of others. Thus, when we think of our own ideal life (or perhaps the life of our favorite historical or public figure), we are often practically right to focus on this life's *roles* for others, and not only, or even mostly, on how it feels from the inside.

---

[153] More in 6.4.1. For a book-length exploration of compassionate impact as a positive purpose in life, see Vinding, forthcoming.

In particular, since the value of the life's roles is ultimately measured in the same unit of value as its own welfare, we can directly say that the roles can be much bigger than what any single life independently "contains".[154]

In the case of minimalist views, we may find that a sufficient reason to endure hardship is to prevent hardship, reduce inner conflict, and lighten the load for all sentient beings over all time.

And if we further zoom into the nature of "effort", we may find that basically all of our daily struggles are much easier to bear compared to instances of the most intense pains.[155] Thus, we may already find some lightness and relief in being relatively problem-free at the personal level. And we may further realize that we can play highly worthwhile roles by focusing our spare efforts on helping to relieve such extreme burdens in the big picture.

By contrast, if we assume that our burdens are worthwhile for the sake of some intrinsic positive value, then we again face theoretical tradeoffs like the VRC, as well as the practical question of whether we would allow astronomical amounts of extreme suffering to take place for the sake of creating astronomical amounts of purported positive goods.

Finally, we might question the practical relevance of thinking that a life could be worth living only for some kind of "self-contained" satisfaction. After all, our practical intuitions and dilemmas never concern tradeoffs between fully self-contained lives, which none of us ever are.

Even without the concept of intrinsic positive value, a life can be worth living for its positive roles.

---

[154] More in 6.2.

[155] Cf. Gómez-Emilsson, 2019; Gómez-Emilsson & Percy, 2023.

Chapter 4

Minimalist Extended Very Repugnant Conclusions Are the Least Repugnant

Population axiology matters greatly for our priorities. Recently, it has been claimed that all plausible axiological views imply certain "very repugnant conclusions" (defined below in 4.1 and 4.2). In this response, I argue that minimalist views avoid these "very repugnant conclusions", and that they face less repugnant conclusions than do offsetting views (4.4).

4.1 Are Repugnant Implications Inevitable?

In population axiology, certain offsetting views, according to which independent bads can be offset by a sufficient amount of independent goods, face the **Very Repugnant Conclusion** (**VRC**):

> A population of arbitrarily many lives with arbitrarily high welfare is worse than a population of arbitrarily many arbitrarily negative lives plus sufficiently many "ε-lives"[156] that each have an arbitrarily small quantity of positive welfare (Figure 4.1).

'worse' outcome　　　　**'better' outcome**

arbitrarily
many
arbitrarily
negative
lives

sufficiently many
lives each with an
arbitrarily small
quantity (ε) of
positive welfare

+

arbitrarily
many lives
with
arbitrarily
high welfare
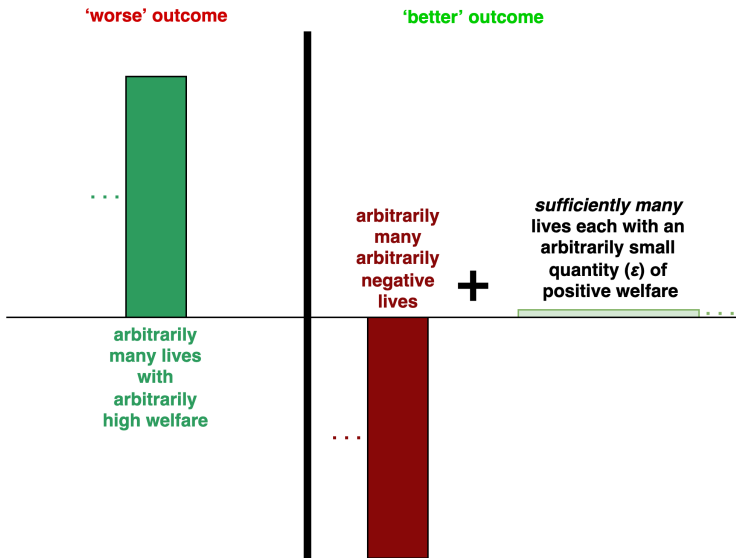
Figure 4.1. The original VRC.

Additionally, offsetting views allow the ε-lives in the VRC to be "rollercoaster lives" that all contain unbearable suffering, purportedly counterbalanced by a sufficient amount of bliss.[157]

In particular, symmetric classical utilitarianism implies interchangeability between a non-suffering ε-life and the rollercoaster life illustrated in Figure 4.2, provided that the "overall welfare" of the rollercoaster life equals ε.

---

[156] en.wikipedia.org/wiki/Epsilon: "In mathematics (particularly calculus), an infinitesimally small positive quantity is commonly denoted ε."

[157] For more on rollercoaster lives, see Appendix 4.
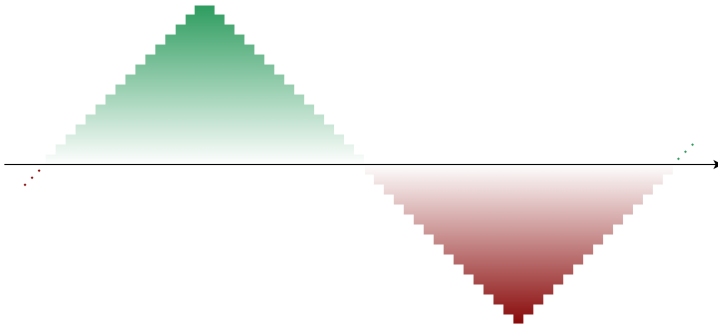
Figure 4.2. The happiness and suffering over time of a single roller-coaster life.

Moreover, many offsetting views, including symmetric classical utilitarianism, would allow replacing each non-suffering ε-life in the original VRC with an "intrapersonal VRC life" (Figure 4.3).
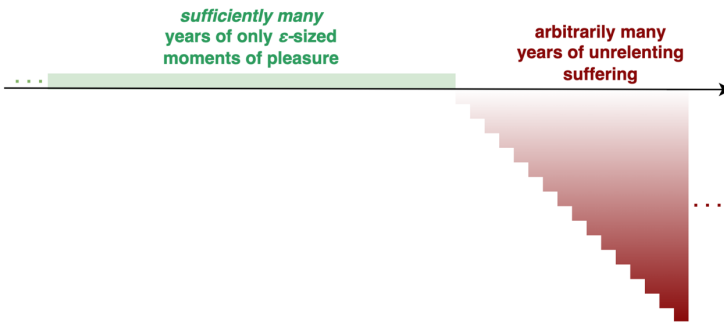
Figure 4.3. The happiness and suffering over time of an intrapersonal VRC life.

Recently, Budolfson and Spears (2018) have argued that all plausible views in population ethics imply similarly repugnant conclusions, namely that they imply either the **VRC** or a closely analogous **Extended VRC** (**XVRC**), which I illustrate shortly at the beginning of 4.2.1.

The purpose of this chapter is to argue that this claim does not apply to minimalist views. In a nutshell: minimalist views avoid

the VRC, can avoid repugnant XVRCs, and, at any rate, face XVRCs that are less repugnant than are the comparable conclusions faced by offsetting views.

4.1.1 Three Claims

Budolfson and Spears make the following three claims:[158]

> Claim 1: No leading welfarist axiology can avoid the **VRC**.
>
> Claim 2: No other welfarist axiology in the literature can avoid the **XVRC**.[159]
>
> Claim 3: The XVRC is just as repugnant as the VRC.

The authors conclude that:

> Repugnant implications are an inevitable feature of any plausible axiology. If repugnance cannot be avoided, then it should not be. We believe this should be among the guiding insights for the next generation of work in value theory.

Below are my brief, initial responses to each of those three claims, and a brief overview of the rest of my response.

4.1.2 Claim 1 Does Not Apply to Minimalist Axiologies

The scope of Claim 1 ("No leading welfarist axiology can avoid the VRC") is limited to 'leading' welfarist axiologies, namely to

---

[158] Budolfson & Spears, 2018, pp. 31–32. (The paper is a precursor to Spears & Budolfson, 2021; I discuss the former because it is more open-access than the latter.)

[159] For an exact definition of the XVRC, see Appendix 4. (For the main text, I will use more intuitive and less formalized descriptions of the XVRC, including some extended variants of the XVRC.)

views that, according to the authors, are commonly-held in the axiological literature.[160]

Thus, the scope of Claim 1 does not cover minimalist axiologies, although axiologies that are essentially minimalist have been defended in the philosophical literature.[161]

To the extent that the VRC seems repugnant, it is worth noting that all minimalist axiologies do avoid the VRC (cf. 3.3.3), and can do so on a principled basis without relying on arbitrary or ad hoc assumptions.

4.1.3 Claim 2 Requires That We Extend the XVRC

Claim 2 ("No other welfarist axiology in the literature can avoid the XVRC") is not straightforward to evaluate, because the original XVRC, as the authors define it, applies strictly only to views that make the assumption of independently aggregable positive utility.[162]

This assumption is not made by minimalist welfarist axiologies, such as antifrustrationism, tranquilism, and some types of negative utilitarianism.[163]

Yet we can slightly extend the original definition of the XVRC, and thereby construct XVRCs for minimalist views. This is done in 4.2.

4.1.4 Claim 3 Requires Comparisons

Finally, we will evaluate Claim 3 ("The XVRC is just as repugnant as the VRC") in the case of minimalist views. If Claim 3 were

---

[160] Budolfson & Spears, 2018, p. 8.

[161] See, for instance, Schopenhauer (1819, "the negative nature of all satisfaction"; 1851, "negative in its character"; Fox, 2022, "pleasures of distraction"), Wolf (1996, 1997, 2004), Fehige (1998), Breyer (2015), and Knutsson (2021b, "axiological claim").

[162] Cf. Appendix 4.

[163] Fehige, 1998; Gloor, 2017; Anonymous, 2015; Vinding, 2022f.

true not only for offsetting views but also for minimalist views, that would support the authors' conclusion that repugnant implications are inevitable.

Yet is it true? That is, are minimalist XVRCs just as repugnant as the VRC?

Settling this question requires that we directly compare minimalist XVRCs against the VRC.

4.1.5 Chapter Overview

Section 4.2 illustrates comparable XVRCs for offsetting and minimalist views.

The illustrations are divided into three separate categories, based on the underlying assumptions about aggregation:

1. Archimedean views ("Quantity Can Always Substitute for Quality") (4.2.1),

2. lexical views with sharp thresholds (4.2.2.1), and

3. lexical views without sharp thresholds (4.2.2.2).

Section 4.3 unpacks which sources of repugnance are present in the different XVRCs, and why I exclude the element of non-creation as non-repugnant.

Section 4.4 evaluates the comparative repugnance of the offsetting and minimalist XVRCs within each of the three categories of views.[164] Additionally, it explains how at least some minimalist views only face XVRCs that are less repugnant than the original VRC, which implies that Claim 3 ("The XVRC is just as repugnant as the VRC") does not hold for those minimalist views.

---

[164] In this chapter, I do not focus on the relative plausibility of different views *across* the three categories of views on aggregation, because my main focus is on the relative plausibility of offsetting versus minimalist views in general, regardless of one's theory of aggregation.

4.2 Comparable XVRCs for Offsetting and Minimalist Views

4.2.1 Archimedean Views ("Quantity Can Always Substitute for Quality")

Let us look at comparable XVRCs for Archimedean views. (Archimedean views roughly say that "quantity can always substitute for quality", such that, for example, a sufficient number of minor pains can always be added up to be worse than a single instance of extreme pain.[165])

Figure 4.4 illustrates the **original XVRC** for Archimedean offsetting views, which goes roughly like this:[166]

---

[165] Thornley, 2022.

[166] Cf. Budolfson & Spears, 2018, p. 19.

> Rather than adding arbitrarily many lives with arbitrarily high welfare, it is better to add arbitrarily many arbitrarily negative lives and have each life in a sufficiently large base population receive an arbitrarily small quantity (ε) of **positive welfare** (an ε-change).
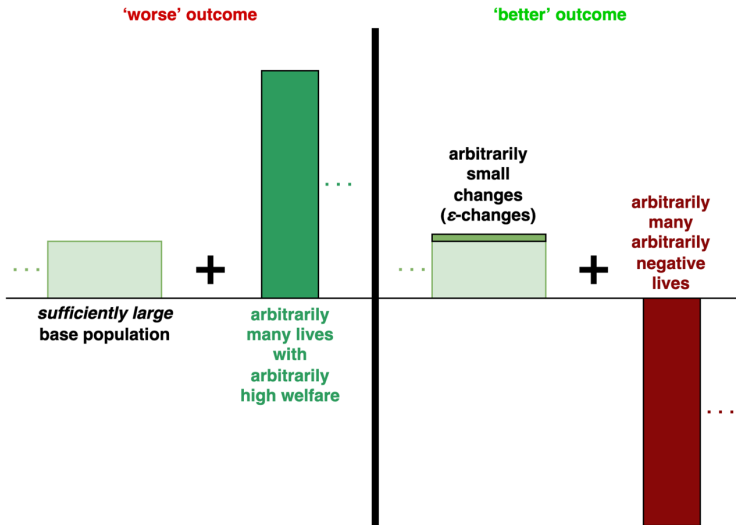
Figure 4.4. An XVRC for Archimedean offsetting views.

A special case of the Archimedean offsetting XVRC is "Creating hell to please the blissful" (Figure 4.5), in which every life in the base population is brought from a very high welfare to an even higher welfare at the cost of adding maximally bad lives.[167]

---

[167] From Vinding, 2021, sec. 3.
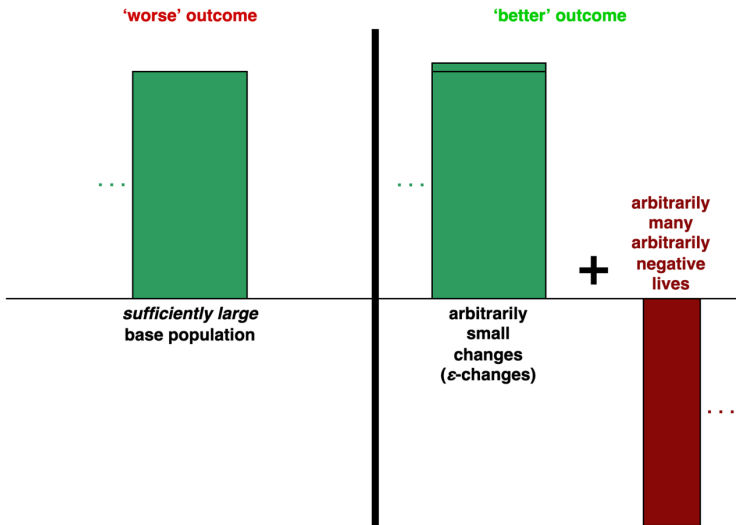
Figure 4.5. Another XVRC for Archimedean offsetting views ("Creating hell to please the blissful").

In the case of minimalist welfarist axiologies, 'welfare' cannot refer to independently aggregable **positive utility**. Instead, minimalist views construe welfare as the absence of intrinsically problematic features, such as 'frustration', 'craving', or 'discontentment'.[168]

Yet we can nonetheless construct an XVRC for Archimedean minimalist views by defining the arbitrarily small changes (ε-changes) more generally as ε-sized **improvements** in welfare.

Thus, an Archimedean **minimalist XVRC** could go like this (Figure 4.6):

---

[168] See, for instance, Fehige, 1998; Anonymous, 2015, sec. 2.2; Gloor, 2017, sec. 2.2. See also Knutsson, 2022b, and Vinding, 2022e.

> It is a net improvement to add arbitrarily many arbitrarily negative lives so as to barely reduce the suffering of each life in a sufficiently large base population.[169]
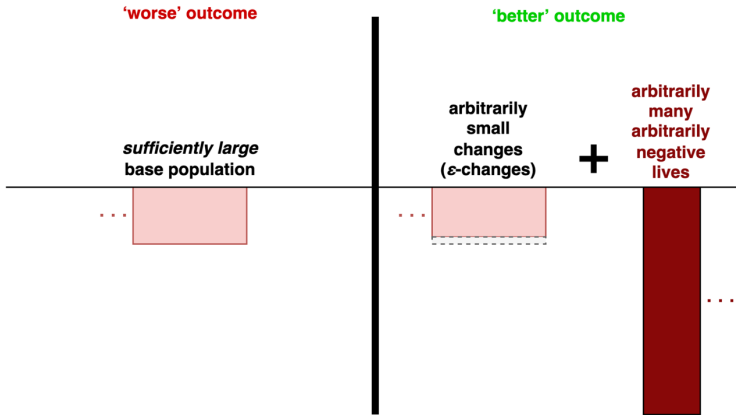
Figure 4.6. An XVRC for Archimedean minimalist views.

Another XVRC for Archimedean minimalist views is illustrated in Figure 4.7, which is basically what is known as the Reverse Repugnant Conclusion.[170]

---

[169] Thanks to Michael St. Jules for pointing out what can be seen as XVRCs for minimalist views.

[170] Cf. Carlson, 1998, p. 297; Mulgan, 2002. This is also known as the Mirrored Repugnant Conclusion, and is analogous to the much-discussed case of "Torture vs. dust specks" (Yudkowsky, 2007).
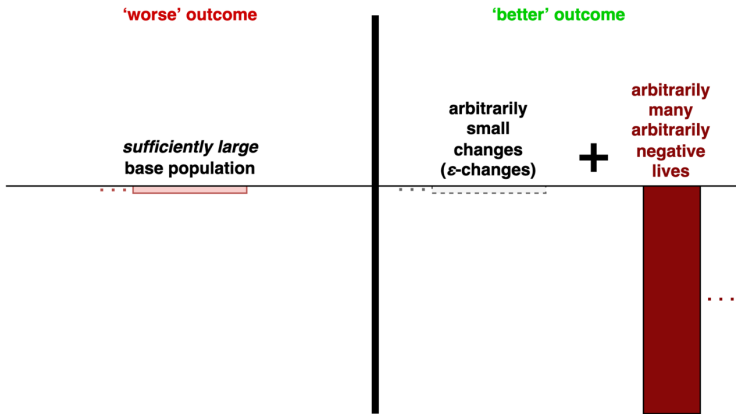
Figure 4.7. Another XVRC for Archimedean minimalist views (the Reverse Repugnant Conclusion).

4.2.2 Lexical Views ("Some Qualities Get Categorical Priority")

Let us now look at comparable XVRCs within a prominent class of non-Archimedean views, namely what are known as **lexical** views. Lexical views deny that "quantity can always substitute for quality"; instead, they assign categorical priority to some qualities relative to others.[171]

Specifically, lexical **minimalist** views entail lexicality between bads, such as by (all else equal) prioritizing the reduction of unbearable suffering over any mild discomfort.[172] Additionally, lexical **offsetting** views entail lexicality between goods (e.g.

---

[171] For an introduction to value lexicality, see Knutsson, 2016c. "Lexicographic preferences" seem named after the logic of alphabetical ordering, in which the "value entities" with top priority are prioritized first regardless of how many others there are in the "queue"; cf. en.wikipedia.org/wiki/Lexicographic_preferences#Etymology.

[172] Knutsson, 2016c, "Lexicality between bads". Cf. Vinding, 2022c.

higher pleasures over lower pleasures), or between goods and bads (e.g. higher pleasures over mild discomfort).[173]

Do lexical views face XVRCs? Notably, lexical views may question the assumption of representing welfare with a single real number to begin with.[174] Thus, lexical views may reject the formal framework of the Archimedean XVRC, whose definition entails arbitrarily small changes (ε-changes) to the aggregate welfare (a real number) of each life in the base population.

However, even if we reject the Archimedean framework, we can still reinterpret the XVRC to construct analogous **lexical XVRCs** (for both minimalist and offsetting lexical views).

Let us first look at such XVRCs for lexical views with sharp thresholds, and then for lexical views without sharp thresholds.

4.2.2.1 With Sharp Thresholds

Consider a lexical **minimalist** view with a sharp threshold. For instance, one may hold that some sentient minds have a sharp breaking point at which suffering becomes unbearable, and that the passing of this point is categorically worth avoiding more than any amount of "bending without breaking".[175]

Figure 4.8 illustrates an XVRC for such a view:

---

[173] Knutsson, 2016c, "Lexicality between goods". A formalism of "higher values" over "lower values" is considered in, for instance, Carlson, 2007; Thomas, 2018; Nebel, 2022.

[174] Vinding, 2022c, sec. 4.

[175] More concretely, the breaking point may be equated with a supposed point at which the suffering becomes "unconsentable" (cf. Tomasik, 2015a, "Consent-based negative utilitarianism?"). For purely minimalist views, one could imagine that this corresponds to suffering so bad that an altruistic agent cannot consent to it even for preventing similar suffering for others. See also Leighton, 2023, chap. 7, "Unbearable Suffering as an Ethical Tipping Point".

It is a net improvement to add arbitrarily many non-lexically bad states (of e.g. barely bearable suffering) as long as we reduce the number of lexically bad states (of e.g. unbearable suffering).
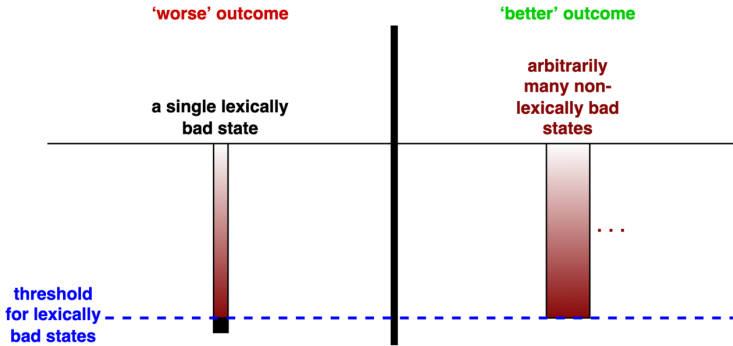
Figure 4.8. An XVRC for lexical minimalist views with a sharp threshold.

A comparable lexical **offsetting** view might entail all of the following claims:

1. There is a lexical threshold between some goods (e.g. higher and lower pleasures).

2. There is a lexical threshold between some bads (e.g. bearable and unbearable suffering).

3. No amount of non-lexical goods (e.g. lower pleasures) can counterbalance a lexically bad state.

4. Some amount of lexical goods (e.g. higher pleasures) can counterbalance a lexically bad state.

Figure 4.9 illustrates an XVRC for such a view:

> It is a net improvement to add arbitrarily many arbitrarily negative lives (that entail lexically bad states) so as to replace, within each life in a sufficiently large base population, a just barely not lexically good state with a lexically good state.
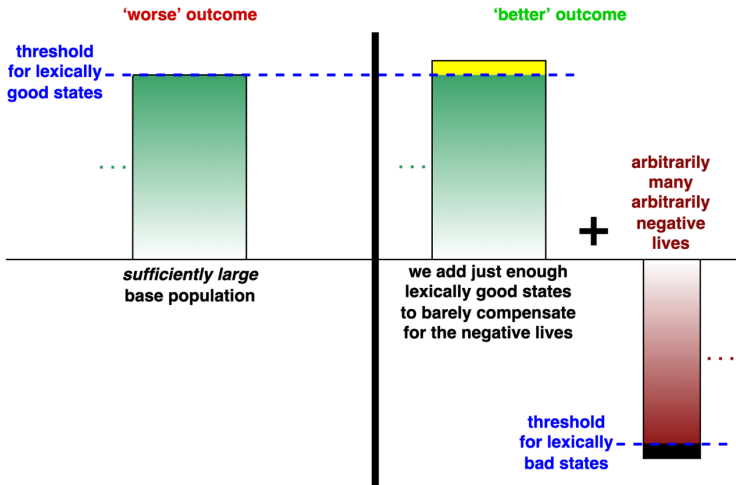
Figure 4.9. An XVRC for lexical offsetting views with sharp thresholds.[176]

4.2.2.2 Without Sharp Thresholds

Finally, it has been argued that lexical views need not entail sharp thresholds like the ones that were abstractly sketched in the previous subsection. After all, perhaps a more plausible lexical view would hold that (e.g.) unbearableness comes in degrees.[177]

---

[176] A way to make each of these lexical XVRCs entail the "arbitrarily small difference" (ε) element of the original XVRC is to make them **probabilistic**, so that the lexically bad state in Figure 4.8, and the lexically good states in Figure 4.9, would happen only with probability ε (cf. Budolfson & Spears, 2018, pp. 12–14).

[177] Cf. Vinding, 2020a, sec. 3. Examples of non-sharp lexical views are presented in Vinding, 2020b, 2022c, and Knutsson, 2021a.

A "non-sharp" lexical threshold could be a **range** (e.g. in the intensity of suffering) between which the suffering becomes lexically worse than suffering above the range. This would imply that no duration of suffering above the range (e.g. "wholly bearable suffering") can be worse than a single instance of suffering below the range (e.g. "wholly unbearable suffering").[178]

Figure 4.10 illustrates a non-sharp lexical **minimalist** XVRC:

It is a net improvement to add arbitrarily many non-lexically bad states (e.g. wholly bearable suffering) as long as we reduce the number of lexically bad states (e.g. wholly unbearable suffering).

'worse' outcome   'better' outcome

a single lexically   arbitrarily many non-
bad state   lexically bad states

range
between
which
lexicality
kicks in
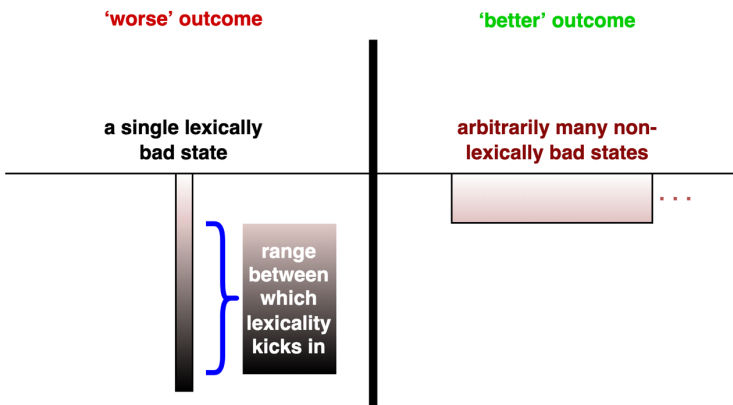
Figure 4.10. An XVRC for lexical minimalist views without sharp thresholds.

---

[178] At the same time, inter-intensity comparisons that take place entirely *within* this range could follow some Archimedean theory of aggregation. Note also that the "non-sharp" lexical XVRCs below are illustrated using only a single range, yet such views could just as well entail multiple different ranges (Vinding, 2022c, sec. 2). (Many interesting details about lexical views are best set aside here, because they apply to both minimalist and offsetting views, and hence those details have limited relevance for my goal of comparing these two classes of views.)

Figure 4.11 illustrates a non-sharp lexical **offsetting** XVRC (based on the same four assumptions that were made for the previous offsetting view).

> It is a net improvement to add an arbitrarily large number of arbitrarily negative lives (that entail lexically bad states) so as to add one lexically good state to each life in a sufficiently large base population.
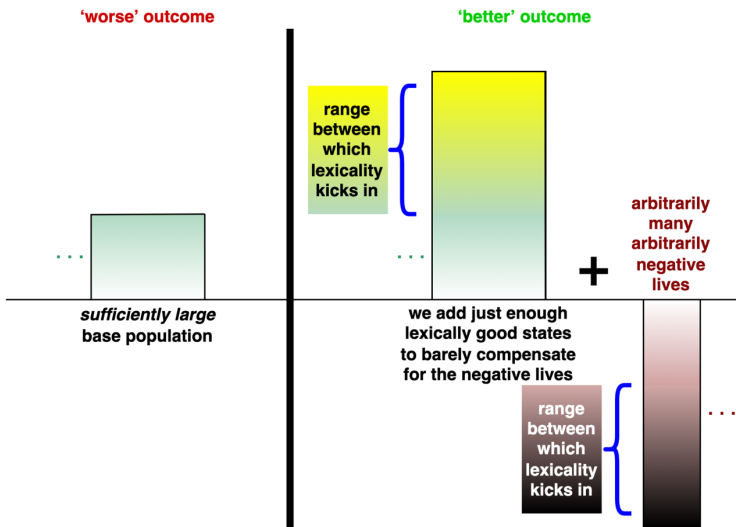
Figure 4.11. An XVRC for lexical offsetting views without sharp thresholds.

4.3 Sources of Repugnance

Section 4.4 will evaluate the comparative repugnance of the above XVRCs and the original VRC. Before that, let us unpack which sources of repugnance are present in the different XVRCs.

4.3.1 Creating Non-Relieving Goods for Some at the Price of Unbearable Suffering for Others

Only the offsetting XVRCs entail the creation of non-relieving goods for some at the price of unbearable suffering for others.

By contrast, the changes in the minimalist XVRCs seem qualitatively less frivolous, because they are about the reduction of suffering rather than about the increase of non-relieving pleasure — pleasure that has no positive roles for relieving anyone's burden.[179]

4.3.2 Enabling Rollercoaster Lives That All Contain Unbearable Suffering

Only offsetting views allow replacing the lives in the various XVRCs with "rollercoaster lives" that all contain unbearable suffering (Figure 4.2). Moreover, Archimedean offsetting views allow replacing them with "intrapersonal VRC lives" (Figure 4.3). In principle, each of these "rollercoaster lives" and "intrapersonal VRC lives" could contain arbitrarily many instances of extreme suffering.

By contrast, minimalist views reject the offsetting ('good minus bad') view of aggregation that enables the rollercoaster interpretation in the first place. Thus, minimalist views entail only the non-rollercoaster versions of the minimalist XVRCs above, whereas the offsetting views entail those same conclusions *plus* their rollercoaster versions.

---

[179] Cf. Vinding, 2020c; 2020d, chap. 3.

4.3.3 Making Seemingly Trivial Changes at the Price of Unbearable Suffering

Budolfson and Spears (2018) seem to attribute repugnance in large part to the unbounded aggregation of arbitrarily tiny changes.[180]

Such aggregation is indeed a potential source of repugnance in at least the Archimedean XVRCs, for both offsetting and minimalist views.

Additionally, perhaps an "atom" of a lexical offsetting good — such as the briefest possible experience of a non-relieving higher pleasure (the authors mention a tiny duration of Mozart) — is intuitively still roughly as trivial as is any other instance of a non-relieving good.

However, one could reasonably argue that the following XVRC statements, made by Budolfson and Spears in the quotes below, do not apply in the case of lexical minimalist axiologies:

> Either something that seems important will be outweighed by an unbounded number of initially unimportant-seeming matters, something that initially seems unimportant will unduly shape the outcome, or both. (pp. 30–31)

This statement does not seem to hold for lexical minimalist views. After all, they would not allow "something that seems important" (such as extreme suffering) to be outweighed by unimportant-seeming matters, nor would they allow "something that

---

[180] Budolfson and Spears (2018): "We note that a common theme emerges, which is that any axiology that aggregates over unbounded spaces will have repugnant implications. This is the fundamental mechanism that our proofs exploit." (p. 2) "Because all plausible axiologies permit aggregation over unbounded spaces, this means that all plausible axiologies are exposed to repugnant conclusions" (p. 30).

initially seems unimportant" (such as non-relieving goods) to unduly shape the outcome. And in particular, extreme suffering seems important, even if it only lasts for a short duration.[181]

> [Our argument] considers the possibility that better-off people are qualitatively different [and] that higher pleasures are qualitatively different. In general, because lexical views still must aggregate across people, they remain subject to repugnance. (p. 34)

As noted, perhaps repugnant implications are inevitable for offsetting views even in their lexical versions (cf. adding tiny durations of non-relieving lexical goods at the price of unbearable suffering).

Yet the authors do not seem to discuss lexical views that give overriding priority to the prevention of extreme bads. Such views are arguably uniquely resistant to this "trivial changes" source of repugnance.[182]

4.3.4 Non-Creation?

The three sources of repugnance covered in the last three subsections each entail the increase of unbearable suffering for the sake of changes that seem relatively frivolous or trivial in comparison.

The original VRC and XVRC additionally entail what might seem like a fourth source of repugnance, namely the non-creation of high-welfare lives whose existence would purportedly be a

---

[181] Vinding, 2020d, sec. 8.12.

[182] It seems plausible to prioritize the reduction of *certainly unbearable* suffering over *certainly bearable* suffering (and over the creation of non-relieving goods) in theory. Additionally, such a priority is, at the practical level, quite compatible with an intuitive, continuous view of the expected amount of unbearable suffering that our decisions may influence — a view that takes into account our uncertainty regarding when and where unbearable suffering might occur (Vinding, 2022d, sec. 2.2; 2022i, sec. 4).

great benefit for their own sake. Let us call this the "non-creation of happy isolated lives".

A key thing to note when evaluating non-creation as a potential source of repugnance is that we need to carefully isolate our intuitions on this question — that is, on non-creation's independent repugnance, all else equal — from the influence of various factors that are actually external to the question itself. When we have done so, I maintain that non-creation is not an independent source of repugnance.[183]

Let us briefly unpack some of the reasons why non-creation is plausibly non-repugnant:

1. How repugnant is the non-creation of lives that are described as "awesome", "flourishing", or "full of love and accomplishment", *all else equal*?[184] Such framings of the lives in question are quite common, yet they may cause our evaluation to become strongly biased in favor of creation. After all, our practical intuitions easily associate those descriptions with lives that play positive roles for others (even on purely minimalist views of value), whereas standard population axiology counterintuitively requires that we ignore all such roles.[185]

2. One way to properly respect the 'all else equal' assumption is to explicitly highlight that the lives in question are

---

[183] Efforts toward such proper isolation are made in 3.4; 5.2.2; and the thought experiment in Vinding, 2022k.

[184] Compare, for instance, Vinding, 2022b, "full of love and accomplishment".

[185] Additionally, our evaluation may be affected by factors such as what we ourselves would like to witness in the world (such as some kinds of inspiring, beautiful, or epic lives), or by us feeling that even a theoretical acceptance of non-creation would have some undesirable implications for our lives. These, too, are subtle ways of breaking the 'all else equal' assumption, because our evaluation of the prospective lives (for their own sake) should be unaffected by the positive roles that the creation or existence of those lives could have for others, including for us (5.2.4).

forever causally isolated lives that never affect any other beings in any way (e.g. that they are happy isolated matrix lives, dwellers of their closed island worlds, or the like, which clearly make no difference beyond themselves). And to further remove our potentially self-related concerns from the picture, we should imagine that no one will know whether we endorsed the creation or non-creation of those happy isolated lives, and that even we will have our memory wiped of the decision immediately after we make it.[186]

3. For experientialist consequentialists, the question of whether non-creation is repugnant becomes a question that arguably requires a thorough phenomenological search, namely a search for non-relieving goods that constitute a positive counterpart to suffering. Yet such a counterpart plausibly does not exist.[187]

---

[186] If the choice starts to feel different when we exclude these factors, this may suggest that the initial framings did indeed evoke factors whose influence was supposed to be ruled out. And this would not be surprising, as our practical intuitions are arguably not adapted to track only the subjective contents of lives, but also (and perhaps even mostly) their overall effects on others. Notably, even the standard question of whether it is "good to create happy people" may still cause a bias in favor of creation, because it may strongly evoke the relational value that these happy, intuitively *prosocial* people would contribute via being good friends, partners, caregivers, citizens, etc. — value whose mental exclusion may be "easier thought than done" (cf. Chapter 3).

[187] Cf. Chapter 1; Gloor, 2017, sec. 2.1; Sherman, 2017, pp. 103–107. A key point here is to not confuse better states with intrinsically positive states, given that our seemingly positive states can be understood as merely better states along a continuum that ranges from states of unbearable torment to states in which we would be completely tranquil or undisturbed. And one may reasonably argue that "intrinsically positive states" could only be reliably identified from a (perhaps very rare) completely untroubled state to begin with (Knutsson, 2022b, sec. 5.2). Finally, we may have reasonable alternative explanations for why we might commonly believe in a positive counterpart to suffering (cf. 1.2; Vinding, 2022e, sec. 4).

4. For preference-based views, it likewise makes sense to think of preference satisfaction as an inherently asymmetric endeavor. Singer (1980): "The creation of preferences which we then satisfy gains us nothing. We can think of the creation of the unsatisfied preferences as putting a debit in the moral ledger which satisfying them merely cancels out."[188] Fehige (1998): "We have obligations to make preferrers satisfied, but no obligations to make satisfied preferrers. … Maximizers of preference satisfaction should instead call themselves minimizers of preference frustration."[189]

4.4 Comparative Repugnance

Let us now evaluate the comparative repugnance of the XVRCs and the original VRC.

Tables 4.1–4.3 show how the minimalist XVRCs explored above are a proper subset of the XVRCs that are implied by the corresponding offsetting views explored above.

Specifically, the offsetting views entail the minimalist implications, their "rollercoaster" or "intrapersonal VRC" versions,[190] and the offsetting implications.

---

[188] It is worth noting that Singer in his article (1980) wrote favorably of combining Preference Utilitarianism and Classical Utilitarianism. Moreover, Singer appears to have moved further toward Classical Utilitarianism in recent years (see e.g. Lazari-Radek & Singer, 2017, chap. 3).

[189] See also DiGiovanni, 2021b. As with experientialist views, we need to isolate our intuitions about the supposedly independent value of preference satisfaction, 'all else equal', from the potentially distorting influence of our practical intuitions. This is because the satisfaction of any given preference in practice often has implications for the satisfaction of other preferences that we may intuitively care about.

[190] Offsetting views also entail the "rollercoaster" or "intrapersonal VRC" versions of the original Repugnant Conclusion (RC) that was discussed in 3.3. While many people may think that the RC is "not necessarily unacceptable" (cf. Zuber et al., 2021), its repugnance can differ greatly depending on whether we allow the population of lives "barely

Notably, only the offsetting implications entail all three of the sources of repugnance that were considered above (that is, creating non-relieving goods at the price of others' suffering; roller-coaster lives; and seemingly trivial changes at the price of unbearable suffering). In contrast, at least some of the minimalist XVRCs entail none of these sources of repugnance.

For the reasons explained in the previous section, I do not count non-creation as a source of repugnance below.

---

worth living" to consist of rollercoaster lives (or intrapersonal VRC lives) that all contain unbearable suffering, compared to if they do not suffer at all.

Table 4.1. Implications of the Archimedean views (4.2.1).

Implication	View Archimedean Offsetting	Minimalist
Very Repugnant Conclusion (Fig 4.1)	x	
rollercoaster version	x	
intrapersonal VRC version	x	
Archimedean offsetting XVRC (Fig 4.4)	x	
rollercoaster version	x	
intrapersonal VRC version	x	
"Creating hell to please the blissful" (Fig 4.5)	x	
rollercoaster version	x	
intrapersonal VRC version	x	
Archimedean minimalist XVRC (Fig 4.6)	x	x
rollercoaster version	x	
intrapersonal VRC version	x	
Reverse Repugnant Conclusion (Fig 4.7)	x	x
rollercoaster version	x	
intrapersonal VRC version	x	

Among the Archimedean implications, the offsetting XVRCs (Figures 4.4 and 4.5) entail more sources of repugnance than do the minimalist XVRCs (Figures 4.6 and 4.7). Specifically, the offsetting XVRCs entail all the same sources of repugnance as does the original VRC, with no redeeming "anti-repugnant" elements.

We can thus agree with Budolfson and Spears' (2018) claim that these XVRCs are just as repugnant as the VRC (i.e. Claim 3).

By contrast, the minimalist XVRCs lack the repugnance of creating non-relieving goods for some at the price of unbearable suffering for others, and the repugnance of "rollercoaster" or "intrapersonal VRC" lives that all contain unbearable suffering.

Thus, the repugnance of the minimalist XVRCs relative to the VRC depends on whether they entail any additional sources of repugnance that the VRC does not. And one may reasonably hold that they do not, which would imply that Claim 3 does not hold for Archimedean minimalist views.[191]

Table 4.2. Implications of the lexical views with sharp thresholds (4.2.2.1).

View	Lexical sharp	
Implication	Offsetting	Minimalist
Lexical sharp minimalist XVRC (Fig 4.8)	x	x
rollercoaster version	x	
Lexical sharp offsetting XVRC (Fig 4.9)	x	
rollercoaster version	x	

Among the sharp lexical views, the offsetting XVRC (Figure 4.9) again seems to entail all the same sources of repugnance as does the original VRC. Only the offsetting view can accept an increase in the number of lexically bad states, such as unbearable suffering. And it does so for the sake of producing non-relieving goods.

---

[191] Of course, one may still find the Archimedean minimalist XVRCs quite repugnant (even if less so than the VRC). Yet the remaining repugnant element of "Making seemingly trivial changes at the price of unbearable suffering" (4.3.3) is avoided by the arguably less repugnant lexical views that give priority to the reduction of lexically bad states.

Table 4.3. Implications of the lexical views without sharp thresholds (4.2.2.2).

View	Lexical non-sharp	
Implication	Offsetting	Minimalist
Lexical non-sharp minimalist XVRC (Fig 4.10)	x	x
rollercoaster version	x	
Lexical non-sharp offsetting XVRC (Fig 4.11)	x	
rollercoaster version	x	

Finally, the non-sharp lexical minimalist XVRC (Figure 4.10) entails that no amount of sufficiently mild states of suffering can be worse than unbearable suffering.

By comparison, the corresponding offsetting view implies not only the same theoretical conclusion (Figure 4.10), but also the arbitrary increase of unbearable states for the sake of creating purportedly sufficient amounts of non-relieving higher goods (Figure 4.11). Additionally, the offsetting view again entails the rollercoaster versions that would "require everyone in the chosen population to experience [arbitrarily] terrible suffering".[192]

---

[192] Budolfson & Spears, 2018, p. 19. The quoted part suggests that the authors do not, in fact, attribute repugnance *only* to the "trivial changes" source of repugnance (4.3.3), but also to the additional suffering entailed by the rollercoaster lives (4.3.2). And if the arbitrary increase of unbearable states (for the sake of additional, unneeded non-relieving goods) is repugnant within lives, then it is presumably just as (if not more) repugnant between lives (4.3.1).

4.5 Conclusions

To respond to the three claims we set out to explore:

> Claim 1: No leading welfarist axiology can avoid the VRC.

Minimalist welfarist axiologies do avoid the VRC. (And they also avoid the "rollercoaster" and "intrapersonal VRC" versions of the original Repugnant Conclusion.)

> Claim 2: No other welfarist axiology in the literature can avoid the XVRC.

Minimalist views do entail modified minimalist XVRCs (cf. Figures 4.6, 4.7, 4.8, and 4.10).

> Claim 3: The XVRC is just as repugnant as the VRC.

This seems true for offsetting views. Yet the minimalist XVRCs are considerably less repugnant than the VRC. In particular, at least the Archimedean minimalist and the non-sharp lexical minimalist XVRCs (Figures 4.6 and 4.10) entail fewer sources of repugnance (4.3) than does the VRC, arguably with no additional, greater sources of repugnance.

Lastly, the XVRCs generated by minimalist views are consistently (4.4) less repugnant than those generated by the corresponding offsetting views. Thus, the minimalist XVRCs seem uniquely unrepugnant. This is a strong point in favor of minimalist views over offsetting views in population axiology, regardless of the specific details of one's theory of aggregation.

Appendix 4

Rollercoaster Lives and the Extended Very Repugnant Conclusion (XVRC)

[This section is entirely quoted from Budolfson and Spears, 2018, pp. 18–19, only annotated with some square brackets by me.]

Let us define an ε-change as a change that makes a difference in either one of these two ways:

> **ε-change:** Let $\varepsilon > 0$ represent any small, positive quantity of well-being. An ε-change either:
>
> - increases the well-being of one person by ε, or
> - adds one new life of well-being ε.

One or more ε-changes can be part of an overall package of changes to a population, but to qualify as an ε-change, a change must be the only change that a particular person receives.

For example, an ε increase could involve slightly improving a tiny headache. One way to see that [an] ε increase could be very repugnant [on offsetting views] is to recall Portmore's [1999] suggestion that ε lives in the restricted RC [Repugnant Conclusion] could be "roller coaster" lives, in which there is much that is wonderful, but also much [terrible] suffering, such that the good ever-

so-slightly outweighs the bad [according to offsetting views]. Here, one admitted possibility is that an ε-change could substantially increase the terrible suffering in a life, and also increase good components; such [an] ε-change is not the only possible ε-change, but it would have the consequence of increasing the total amount of suffering.

With this definition of ε-change in hand, we can now characterize the

Extended very repugnant conclusion (XVRC):

For any:

- Arbitrarily large number of arbitrarily high utility people: $n^h > 0$, $u^h > 0$,

- Arbitrarily large number of arbitrarily negative utility people: $n^l \geq 0$, $u^l < 0$, and

- Arbitrarily small positive quantity of well-being: $\varepsilon > 0$,

There exists:

- A number of ε-changes: n^ε, and

- A (possibly empty) set of base population lives,

such that it is better to both add to the base population the negative-utility lives and cause n^ε ε-changes than to add the high-utility lives.

The XVRC extension from the VRC retains all of the repugnance of choosing many terrible lives over many wonderful lives for merely ε-benefits to other people. Moreover, if ε-changes are of the "roller coaster" form [which would be ruled out by minimalist views], they could increase deep suffering considerably beyond even the arbitrarily many [u < 0] lives, and in fact could require everyone in the chosen population to experience terrible suffering.

Chapter 5

Peacefulness, Nonviolence, and Experientialist Minimalism

5.1 Overview and Scope

This chapter addresses concerns regarding the implications of minimalist views in relation to the following questions:

> **A.** Would an empty world (that is, a world without sentient beings) be axiologically perfect?
>
> **B.** For any hypothetical world, would the best outcome be realized by pressing a button that causes its instant cessation?
>
> **C.** Would minimalist consequentialism imply that it would be right to seek to turn our world into an empty one in practice, even by coercive means?

The scope of my responses is limited to purely welfarist consequentialism.[193] Thus, I assume that the value of outcomes and the rightness of acts depend entirely on wellbeing outcomes, without being affected by any independent disvalue assigned to other factors, such as acts, motives, or character traits.

The scope is further limited to purely experience-focused (that is, 'experientialist') versions of minimalist views. Thus, I assume that the wellbeing of any given being cannot be directly affected by things outside their experience, and thereby set aside extra-experientialist views, such as preference-based views.[194]

Finally, I assume that the independent disvalue of any given experiential state, all else being equal, is wholly intrinsic to that experiential state, independent of the context or the rest of the life in which that state exists.

Extra-experientialist or nonconsequentialist versions of minimalist views could have more things to say about the questions above.[195] On extra-experientialist grounds, such views might say that the outcome resulting from instant cessation could be suboptimal due to other independent bads, such as preference frustration, premature death, or rights violations (i.e. "extra-experiential disvalue").[196] Moreover, nonconsequentialist versions of minimalist views might hold that the rightness of acts, such as pressing a button or engaging in supposedly justified coercion, is dependent also on nonconsequentialist factors, such as some properties of the actor or the action itself, and thereby deny that the outcome is the full picture of what matters morally.[197]

---

[193] en.wikipedia.org/wiki/Welfarism.

[194] An argument for experientialism is, for instance, van der Deijl, 2021. For an overview of minimalist views of wellbeing, see Chapter 1.

[195] See, for instance, Anonymous, 2015; sec. 2.1.1, sec. 2.1.5; Vinding, 2022a, sec. 2.

[196] Cf. 2.3.1.

[197] Cf. 2.2.

Thus, this chapter focuses merely on how the questions listed above could be addressed from a narrow subset of minimalist perspectives. It does not cover all possible replies from minimalist views in general, and hence agreement with the arguments presented in this chapter is by no means required for holding a minimalist view of value and ethics. (That being said, many of the arguments in this chapter, especially the ones that address practical issues, may still be relevant to minimalist views more broadly.)

<div align="center">• • •</div>

The first question (**A**) concerns the theoretical value of hypothetical worlds. Here, experientialist minimalist axiologies do imply that there is no world that could be *better* than an empty world; after all, an empty world would involve no trouble of any kind.[198] Yet an *equally* ideal world, in theory, would be one in which all lives are completely untroubled.[199] And this equal ideal (of completely untroubled experience) is a much more widely shared ideal.

Thus, it is misleading and needlessly divisive to talk as if minimalist views would necessarily prioritize an empty world as their only ideal of perfection — especially after we account for the various practical reasons to focus on the common ground between people who hold different values.

The other questions and related concerns are a lot less trivial. To respond to them in their proper context, I draw a sharp distinction between the hypothetical question (**B**) and the practical question (**C**).

Sections 5.2 and 5.3 address these questions, respectively. The next two subsections provide an outline of the main points.[200]

---

[198] Cf. Knutsson, 2016a.

[199] Pearce, 2007.

[200] This chapter is essentially a two-part response to what is often called the 'world destruction' or 'benevolent world-exploder' argument against 'negative' consequentialist views, framed in less problematic terms.

5.1.1 Overview of the Hypothetical Side

Section 5.2 is a six-part analysis of the hypothetical choice of cessation (**B**).

To focus on the choices in which experientialist minimalist views imply that cessation is the best option, I assume that the equal ideal of an untroubled paradise is not an option. That is, I discuss the non-cessation versus cessation of any given hypothetical world that does contain some amount of experiential bads, such as intense suffering.

The structure of the response is as follows.

- Section 5.2.1 argues that we need to account for status quo bias and omission bias before we could hope to evaluate the choice between non-cessation and cessation from an impartial perspective.

- Section 5.2.2 presents a 'reversal test' to account for those biases, and highlights the need to be mindful of the radical assumption of 'all else being equal' often made in population ethics.

- Section 5.2.3 responds to the choice (between the non-cessation and cessation of any hypothetical world that contains bads) from the perspective of experientialist and consequentialist minimalism.

- Section 5.2.4 explores some implications of our possibly failing to maintain the assumption of 'all else being equal' when engaging in this thought experiment.

- Section 5.2.5 zooms out to ask whether the principles provided by minimalist views (to respond to non-cessation versus cessation) are any less plausible than those provided by offsetting ('good minus bad') views. Additionally, this section:

  - Argues that in comparison with offsetting views, the cessation button does not constitute a unique

        objection against minimalist views in particular (if at all).

- ○ Suggests that offsetting views imply similar or worse hypothetical choices, including the choice of 'intense bliss with hellish cessation'.

- ○ Notes that for some people, these cessation implications may be a reason to reject all purely experientialist and consequentialist views; for others, they may be a reason to highlight the gap between consequentialist theory and practice.

- Section 5.2.6 briefly highlights the gap between consequentialist theory and practice (before 5.3 does so at length). In particular, this section:

- ○ Asks whether our practical anti-violence intuitions (strong and warranted as they are) might "miss their mark" in thought experiments that involve the cessation of causally isolated lives (be it 'minimalist cessation' or 'intense bliss with hellish cessation'), and whether this might constitute an additional bias in such thought experiments.

- ○ Looks at experientialist minimalist reasons to strongly oppose even painless killing in practice.

5.1.2 Overview of the Practical Side

Section 5.3 is a three-part response to the practical question (**C**).

The hypothetical conclusions in Section 5.2 are based on assumptions that are completely unrealistic. In contrast, Section 5.3 explores the altogether different considerations that are of key relevance once we drop those assumptions of 'all else being equal', cessation buttons, and the like.

The structure of the response is as follows.

- Section 5.3.1 acknowledges that we may sometimes practically benefit from orienting toward ideal states-of-affairs (i.e. 'utopianism'). However, it also argues that:

    ○ Utopianism as a guiding principle has many pitfalls that can cause it to diverge from impartial consequentialism (and do more harm than good).

    ○ In practice, consequentialism does not recommend seeking possible paths to a single, ideal "endstate" of the world; instead, a more plausible understanding of consequentialism recommends that we follow indirect proxy principles that yield the best consequences over all space and time in terms of carefully estimated expected value (and do so from a 'marginal realist' rather than a 'broad idealist' perspective).

        ■ The expected value approach is a more impartial, realistic, and risk-aware outcome-orientation that does not privilege any subperiod in time, and nor does it privilege some uncertain prospect of realizing a hypothetical endstate at the risk of leading to an overall worse outcome.

- Section 5.3.2 looks at the key practical considerations for assessing whether minimalist consequentialism, combined with careful expected value thinking, would recommend or discourage efforts to create an empty world. Overall:

    ○ Minimalists have strong practical reasons to cooperate with people who hold different values and to seek mutual gains from compromise with them (especially because the active prevention of worst-case outcomes, a top cause for minimalists, is already a key shared aim between multiple views, and requires us to work together).

○ We still have considerable empirical uncertainty about how common life is in the reachable universe, and about whether an Earth-originating civilization would use astronomical resources in more or less suffering-conducive ways than would another civilization.

- ■ This uncertainty is a reason against prematurely concluding that the extinction of humanity would be desirable compared to human space colonization (even from a minimalist perspective).

- ■ And (as a practically relevant thought experiment), the more we would be both capable and goal-aligned enough to prevent all suffering on Earth, the more we might also be the kind of civilization that could play more positive roles by ensuring that the vast resources of the reachable universe would not become fuel for generating astronomically greater suffering.

○ Minimalists would likely be wise — given the empirical uncertainty — to prioritize the widely shared (and robustly positive) aim of "improving the expected quality of future lives conditional on their existence", such as by working together with people who hold different values to reduce risks of astronomical suffering (**s-risks**).

- ■ S-risk reduction will plausibly require large-scale cooperation between various agents, whereas unilateral action toward an 'empty world' would likely increase s-risks in expectation.

- Section 5.3.3 notes that the aim to minimize experiential bads is not by itself a directly suitable principle for guiding practical action. Rather:

  - Minimalists (and consequentialists in general) need to deeply internalize and uphold more actionable principles — such as virtues and commonsense prohibitions — that indirectly tend to bring about the best consequences.

    - For minimalists, a practically optimal set of principles will likely include pragmatically absolute nonviolence, non-aggression, and respect toward other sentient beings, not least because the erosion of these principles is a prime risk factor for s-risks.

5.2 The Hypothetical Side: Cessation Versus Creation of an Imperfect World

5.2.1 Status Quo Bias and Omission Bias: Privileging Non-Intervention

Status quo bias is an irrational preference for the current situation even when better options are available.[201] Its influence can be probed with the reversal test, which involves hypothetically reversing the current situation to see if we would still prefer it for its own merits, or if our preference for it was influenced by a resistance to change.[202]

For instance, consider Bob, who opposes giving pain relief access to those who lack it (even assuming no switching costs). Would he support removing pain relief access from the same patients if they already had it? If not, then his judgment may be influenced by a preference for maintaining the status quo (or a preference to avoid taking personal action to change it), instead of just assessing each change on its own merits.

Status quo bias may in part be explained by omission bias, namely the tendency to deem harmful inaction (omission) more acceptable than equally harmful action (commission).[203] This, in

---

[201] Cf. en.wikipedia.org/wiki/Status_quo_bias.

[202] Cf. Bostrom & Ord, 2006. For a brief introduction, see forum.effectivealtruism.org/topics/reversal-test.

[203] Cf. en.wikipedia.org/wiki/Omission_bias. For example, say we have two identical populations, P1 and P2, living on identical deserted islands, yet only P1 is afflicted by an infectious disease. We could immunize P1 against this disease at no cost to us. The outcome of our *non-immunization* of P1 is the same as if we were to actively introduce the disease to P2 (a population gets sick). Given these assumptions, if we still tend to consider the non-immunization of P1 more acceptable than actively introducing the disease to P2, this suggests that we may have a bias in favor of non-intervention.

turn, may in part be explained by the fact that we often find it much easier to attribute blame for the harm to someone who took harmful action compared to someone who took no action, even if they both allowed equal harm to take place in the outcome.[204]

Regarding the cessation of a hypothetical world, clearly the choice of cessation would be a significant intervention to the status quo. To account for the influence of the potential biases above, we therefore need a reversal test where the choice that is equivalent to *non*-cessation entails a similar responsibility for the outcome as does the choice of cessation.

In this case, the reversal test is to ask whether we would actively *create* a world that is identical to the one that the choice of non-cessation would passively allow to exist.[205]

5.2.2 The Reversal Test: Creation at the Moment of Non-Cessation

In this context, the reversal test is motivated by the consequentialist equivalence between action and inaction with identical outcomes.[206]

From an experientialist consequentialist perspective, the outcome of the non-cessation choice is that a world *does* exist in place of an empty world, which is outcome-equivalent to the *creation* of a similar world — from the moment of the choice onward — in place of an empty world.

This equivalence is illustrated in Figure 5.1.

---

[204] The 'act-omission distinction' is relevant on many nonconsequentialist views, but it is immaterial on pure consequentialism, which is assumed in this response.

[205] Essentially the same point has often been made by David Pearce. For example: "I could challenge status quo bias and ask critics whether they'd press a notional ON button that generates a type-identical copy of the world – and thereby create more suffering than Adolf Hitler." sentience-research.org/the-imperative-to-abolish-suffering-an-interview-with-david-pearce.

[206] Cf. Woollard & Howard-Snyder, 2022.

time *T* and forward

non-cessation

creation

identical post-*T* outcomes

cessation

non-creation

identical post-*T* outcomes
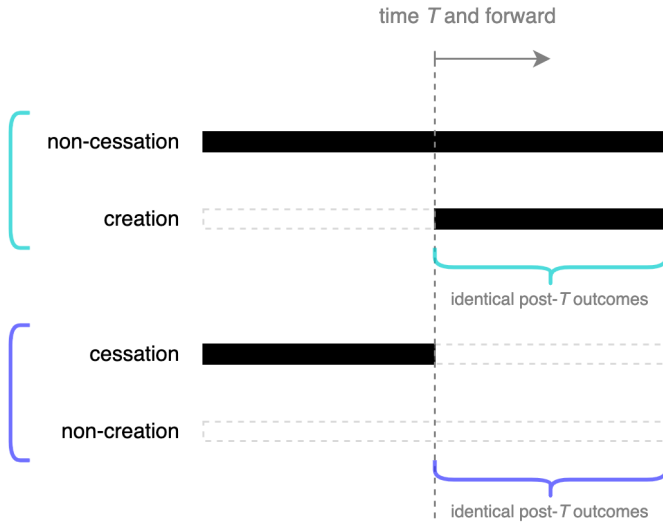
Figure 5.1. The consequentialist equivalence between non-cessation and creation, and between cessation and non-creation.

We can now avoid the influence of status quo bias and omission bias by reframing the initial question (**B**) in more neutral terms, namely as a choice between the middle options in Figure 5.1. Thus, we choose between the creation versus cessation of a hypothetical world *W* at some time *T*:

1. 'Creation': the instant creation of world *W* (starting from time *T*) in place of an otherwise empty world.

2. 'Cessation': the instant cessation of world *W* at time *T*.

From a forward-looking, experientialist consequentialist point of view, the reframed question helps us focus on the main issue of which choice has the overall better outcome.

We, in this thought experiment, are simply an outside 'chooser'. We do not live in world *W*, and are not affected by our choice in any way.

To even better respect the assumption of 'all else being equal' in these hypothetical cases, let us imagine that the entire population of world *W* consists of isolated Matrix-lives that have no ef-

fects beyond themselves.[207] This helps to ensure that our assessment of this hypothetical choice is not distorted by our practical intuitions, which in the real world might implicitly be tracking the overall value of individual lives not only in terms of their independent features, but also in terms of their positive or negative roles for all other lives.[208]

5.2.3 Choosing a Future with Fewer Problems

Now to make the choice, we only need to specify two things:

1. What the future is actually like, in the creation scenario, for the isolated Matrix-lives of world W (from time T onward).

2. What axiological principle[209] we use to assess which choice has the overall better outcome, all else equal.

Trivially, if the lives in world W are completely untroubled, then minimalist views would consider the world an optimal one,

---

[207] Cf. 3.4. To stress-test our intuitions about the independent value of isolated pleasure, we may consider a thought experiment in which we choose between (1) transforming Earth into a 'minimalist paradise' in which everyone is completely untroubled, or (2) ending this world and replacing it with a sufficient number of happy Matrix-lives enjoying maximal bliss. (This is virtually identical to David Pearce's thought experiment about a 'utilitronium shockwave', hedweb.com/social-media/pre2014.html)

[208] Cf. 3.2.3; 3.6. Thus, we need to be careful not to unwittingly misapply our practical intuitions about the cessation versus creation of an individual, real-world life to the hypothetical case of the cessation versus creation of an entire world. After all, an individual, real-world life practically always has relevant effects beyond itself, whereas the set of all lives never does. (A different question, considered on the practical side, is whether Earth-originating life could play positive roles beyond Earth, such as by reducing the suffering caused by other civilizations.)

[209] Axiological as defined in 3.1.

neither favoring its cessation nor opposing its creation. After all, its creation would harm no one.

The more interesting case is the 'near-perfect paradise', in which the isolated Matrix-lives are arbitrarily numerous, blissful, and subjectively meaningful, except one of the lives is rendered 'imperfect' due to a momentary subjective problem. This life would experience some involuntary suffering, unmet need, or preference frustration, of the minimum severity to qualify it as a problem for its own sake.[210]

All else equal, would experientialist minimalist views favor the cessation (i.e. non-creation) rather than the creation of such an 'imperfect world' at time T, before the problem occurs?

Hypothetically and strictly from a consequentialist point of view, the answer is yes. Such views would say that the better future is the one that involves the least amount of involuntary suffering, unmet need, or the like, regardless of how numerous or blissful the other lives are. After all, a core feature of minimalist views is that the suffering, need, or frustration of some beings cannot be counterbalanced or offset by adding subjectively perfect experience-moments elsewhere.[211]

---

[210] Cf. 3.3.4. For this scenario to pertain to minimalist views centered on overall negative states, we need to be careful to not interpret this 'subjective problem' as something that the minimalist views would actually find unproblematic, such as a dip in "non-relieving bliss" from, say, "+100" to "+99". Instead, we should imagine that this episode of slight discomfort would be experienced as an overall negative "−1" episode (Vinding, 2021, "What is the bad in question?").

[211] Cf. Vinding, 2020c.
Among experientialist consequentialists, a seemingly common view of personal identity is empty individualism or "constant replacement", in which we count each "experience-moment" as a separate being (cf. Karnofsky, 2021). This casts doubt on how exactly interpersonal value aggregation or compensation is supposed to work (cf. DiGiovanni, 2021b). I often speak of "beings", "lives", and "experience-moments" interchangeably, and we may interchange them in thought experiments to see whether doing so makes a difference. After all, this substitution should arguably make no difference when we operate within the com-

5.2.4 Minimalist Creation for Extrinsic Reasons: Breaking the 'All Else Equal' Assumption

Minimalist axiologies imply that replacing an empty world with the future of world *W* cannot be an improvement within that world, because the empty world is perfectly fine to begin with ("no need, no problem"). Yet even if one holds this view, one might still intuitively feel that 'creation' would be the better choice in the near-perfect paradise case. This section explores some reasons why such an intuition might arise not necessarily due to our holding a non-minimalist axiology, but possibly due to our breaking the boundaries of the 'all else equal' assumption when engaging in this thought experiment.

First, if we ourselves already experience more subjective trouble contemplating a world's cessation than what trouble would be entailed by its creation, this may in part explain why we intuitively feel that cessation would be the worse choice.[212] For example, we might implicitly feel as if even a hypothetical endorsement of cessation would have dispiriting implications for our own unmet needs for existential security, bliss, or meaning. Yet just because *we* might have such unmet needs, it does not follow that the creation of isolated beings with satisfied such needs would be an independently positive endeavor.

---

bined assumptions of 'all else being equal', experientialism, and consequentialism. So this could be a way to notice the influence of our practical intuitions, which may track things like positive roles even when such factors are supposed to be ruled out.

[212] This point might have been inspired by a related point made by David Pearce (NU here refers to negative utilitarianism): "[Think] of some NU-sounding policy proposal that you find unappealing – or even the slightest bit disappointing to contemplate. Other things being equal, this policy-option can't really be NU … For NUs want to abolish disappointment, frustration and anything that causes you the slightest hint of concern or sadness." sentience-research.org/the-imperative-to-abolish-suffering-an-interview-with-david-pearce

Second, we might be socially aware that even a hypothetical endorsement of emptiness instead of an imperfect world could be misunderstood as an "anti-life" stance in the eyes of others who might care about which "side" we would take if we had to choose between conflicting ideals or policies in practice.[213] Thus, we might feel that the path of least harm is to choose the creation response in the imperfect paradise case, and to hope to convey a more nuanced understanding of our views in deeper discussions, outside the often charged context of unrealistic hypotheticals.

The above considerations are subtle ways of breaking the 'all else equal' assumption, because our hypothetical choice is not supposed to depend on how it might affect *us* in any way. At the same time, we have good reasons to remember that our hypothetical choices never happen in a vacuum. In particular, they can lead to practical misunderstandings, especially given that our responses to them are usually more memorable than the tower of abstract assumptions on which they were based.

To prevent such practical misunderstandings, it is worth noting that the near-perfect paradise *is* as perfect as any world that could ever be practically realized, according to both minimalist and other welfarist views.[214] And as far as utopian ideals are concerned, minimalists have strong, practical reasons to side less with the hypothetical ideal of emptiness or cessation, and more with the equal and more popular ideal of completely peaceful lives.[215]

---

[213] Cf. "The side-taking hypothesis for moral judgment", DeScioli, 2016.

[214] Cf. Tomasik, 2013d.

[215] After all, a shared goal between minimalists and proponents of other welfarist views is to increase the quality of all lives that will exist (other things being equal). By comparison, the ideal of an empty world is not nearly as universal. For more on practical reasons to prioritize common goals between different value systems, see 5.3.2.

5.2.5 Comparative Theoretical Implications of Minimalist and Offsetting Views

This section compares the hypothetical cessation implications of minimalist views with some hypothetical implications of other consequentialist views.

For brevity and ease of reference, the other views will be called "offsetting views" due to their assumption that any independent bads can always be counterbalanced or offset by a sufficient addition of independent goods.[216] Broadly speaking, minimalist axiologies favor outcomes where the notional sum of independent bads is minimized, while offsetting axiologies favor outcomes where the notional sum of independent goods over independent bads is maximized.[217]

Minimalist views are not alone in sometimes favoring the cessation of hypothetical worlds. After all, offsetting views also favor cessation whenever the notional sum of future goods and bads is negative.

Thus, the question is not *whether* offsetting views imply cessation, but rather *when* they imply it. And the question is also what else they imply. In particular, when we compare axiological views with respect to their hypothetical implications, it seems relevant to compare their (apparently) least plausible implications with each other.

Even before looking at the specific implications, one might defend an offsetting view by saying that it favors cessation only in the cases where the "sum" of goods over bads is negative (that is, in the "correct" cases), and that minimalist views would imply cessation in cases where this "sum" ought to be seen as positive (that is, in the "wrong" cases).

---

[216] For an overview of reasons to doubt the offsetting premise, see 1.2.

[217] Offsetting views include, for instance, classical utilitarianism (cf. Sinnott-Armstrong, 2023, sec. 1), as well as "weak negative" or "negative-leaning" views (cf. Tomasik, 2013e; Knutsson, 2016d).

Yet the existence of this "sum" is a huge assumption to begin with.[218] And it is an assumption that can justify arbitrarily severe harms, such as extreme suffering, for the creation of supposedly independent goods, even if the absence of these supposed goods would cause no experiential problem in the first place.[219]

Thus, we may find the offsetting implications much worse than implications that involve no experiential bads, such as choosing the hypothetical scenario of minimalist cessation (that is, universal cessation or non-creation to prevent all future problems[220]).[221]

5.2.5.1 Three Offsetting Implications

Consider, for instance, that offsetting views have the following hypothetical implications:[222]

1. *'Intense Bliss with Hellish Cessation'* (Figure 5.2). Assume a "minimalist paradise" that contains sentient beings experiencing complete peace and contentment, but not (purported) positive pleasure.[223] All else equal, offsetting

---

[218] Knutsson, 2016b; Vinding, 2020c.

[219] Cf. DiGiovanni, 2021b.

[220] Cf. 5.2.3.

[221] Regarding the question of how popular it might be for people to intuitively accept the offsetting premise, compare the three surveys of Tomasik, 2015c, "Pain-pleasure tradeoff"; Future of Life Institute, 2017; and "How many days of bliss to compensate for 1 day of lava-drowning?",
facebook.com/groups/effective.altruists/permalink/1117549958301360.

[222] Thanks to Magnus Vinding for suggesting all three thought experiments. The first one was also partly inspired by Knutsson, 2021b, sec. 3.

[223] In other words, the minimalist paradise only contains sentient beings who spend their entire, eternal lives in subjectively flawless states of tranquility, flow, or the like (cf. Appendix 1; Gloor, 2017, sec. 2.1; Knutsson, 2022b, sec. 2.1), yet who never experience anything that offsetting views would see as positively good, beyond the "mere" complete absence of bads.

views imply that it is better to reject the eternal continuation of this paradise and to instead choose a sufficiently large paradise of intense bliss that ends in an arbitrarily large and hellish cessation.
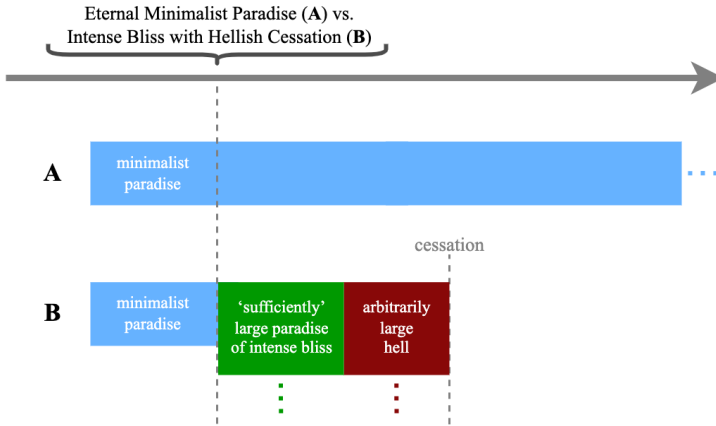
Figure 5.2. An illustration of the choice of 'Intense Bliss with Hellish Cessation' (**B**), which is favored by offsetting views.[224] Minimalist views would favor the continuation of the minimalist paradise (**A**).

2. *'Creating Hell to Please the Blissful'* (Figure 5.3).[225] Say we have a vast population experiencing nearly maximal bliss. All else equal, offsetting views like classical utilitarianism imply that it is a net improvement to add to this population a smaller population of maximally hellish lives, provided that this fully maximizes the bliss of the sufficiently vast population of near-maximally blissful lives.

---

[224] In particular, if we assume "equal intensities" for the supposedly independent goods and bads in the diagram, then classical utilitarianism favors the choice of 'Intense Bliss with Hellish Cessation' (**B**) at the proportional scale that is shown in the diagram. For "negative-leaning" offsetting views to favor **B**, the large paradise would need to extend considerably more than the hell does.

[225] From Vinding, 2021, sec. 3. Also illustrated in Chapter 4. Expressed in hedonistic terms, but applicable to independent goods more broadly.
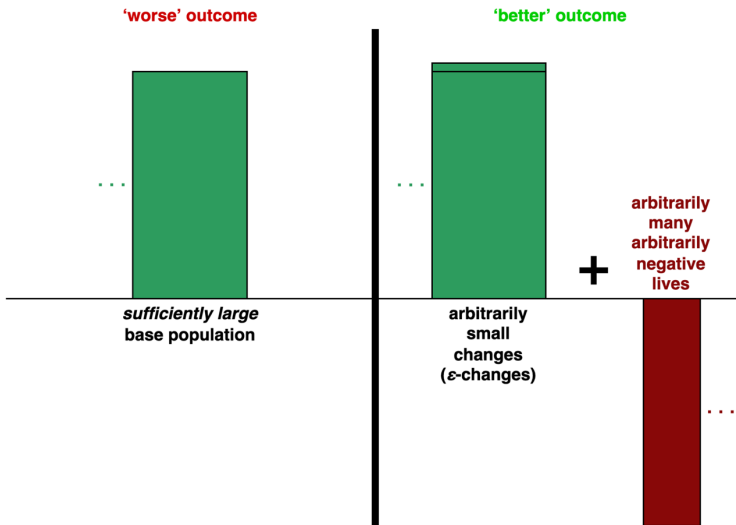
Figure 5.3. 'Creating Hell to Please the Blissful'.

3. *'More Suffering and a Greater "Sum" from Giving Space to an Alien Civilization'.* Assume a world in which humans can either expand into space or give it to an alien civilization.[226] The alien civilization would use the same resources to generate astronomically more suffering, but also vastly more (purported) goods, such that the "offsetting sum" from the alien expansion would be greater than that from human expansion. Here, offsetting views imply that an arbitrarily hellish human extinction is better than human space expansion, as long as the alien civilization generates sufficient goods. By contrast, minimalist views (in this hypothetical) imply that human space expansion is the better outcome, due to ours being the more peaceful and harm-free civilization.

---

[226] Cf. grabbyaliens.com.

5.2.5.2 Cessation Implications Exist for All Experientialist Consequentialist Views

Finally, we may draw an analogy between the case in which minimalist views would favor the cessation of a near-perfect paradise, and the following cases in which *all* experientialist consequentialist views would have a comparable implication.

Namely, just like a single 'near-negligible problem' may lead minimalist views to prefer a hypothetical world's cessation over its creation, so too would a single such problem "tip the scales" according to all experientialist consequentialist views in some worlds.

Specifically, all such views would favor cessation for the following worlds (that are assumed to be "balanced" or "neutral" in terms of experiential value), which may commonsensically seem to contain a lot of value:

1. *'The Pinprick Argument'* (adapted).[227] Regardless of what else a world contains, if the "sum" of all the future experiential goods and bads of its inhabitants were set to be perfectly neutral, *all* experientialist consequentialist views would favor the instant cessation of the world over adding the tiniest of bads.

2. *'Extra-Experiential Fulfillment'*.[228] Say we have a world that contains arbitrarily many lives, all of which are full of knowledge, accomplishments, and relationships, yet where these lives experience none of the supposedly independent goods and bads such as pleasure and pain. According to all experientialist consequentialist views, it would be better to painlessly end these lives than to add the tiniest of bads.

---

[227] Pearce, 2005.

[228] Thanks to Magnus Vinding for this thought experiment.

To commonsense intuitions, these may seem like absurd conclusions, and thereby like compelling reasons to reject all experientialist consequentialist views in general.[229]

However, one could defend all such views in ways that would be notably similar to how we approached the minimalist cessation hypotheticals in the previous sections. That is, one could emphasize that we need to (1) account for our status quo and omission bias, (2) remember to carefully respect the radical assumption of 'all else being equal', and (3) count only that which has *independent* value.

Experientialist offsetting views and experientialist minimalist views both share the approach of reducing seemingly independent values to an underlying intrinsic value, yet the minimalist views in some sense go one step further. That is, experientialist offsetting views already deny that accomplishments, relationships, or any other things have value *independent* of their roles in relation to certain experiential features, such as suffering or (purported) positive pleasure. Experientialist minimalist views do the same, except they take the extra step of denying that any experiences have positive value independent of their roles in relation to troubled experiences.[230]

---

[229] By contrast, when it comes to extra-experientialist versions of consequentialist views, only their offsetting variants can imply that it is a net improvement to create a world with a 'sufficiently' large amount of goods like complex knowledge or accomplishments, no happiness, and arbitrarily many lives full of arbitrarily severe suffering.

[230] (For more on this experientialist minimalist step, see Chapter 1; 3.1; 3.2; 3.5.2; 6.4.) As noted in footnote 49 of DiGiovanni, 2021a, offsetting utilitarians already argue that people may be systematically conflating the instrumental value of non-hedonic goods with their being independently valuable. The minimalist perspective would add that this also holds for what we often call positive experiences. DiGiovanni:

> [T]here is a prima facie argument that strong axiological asymmetries should seem especially plausible to those sympathetic to a hedonistic view. This is because the [offsetting] hedonistic utilitarian already holds that most people are systematically mistaken about the intrinsic value of non-hedonic goods. The

For a person attracted to experientialist and consequentialist views, is the existence of cessation implications itself a reason to reject minimalist views and prefer offsetting views? Clearly not, since we have just illustrated that cessation implications exist for all experientialist consequentialist views.

Moreover, the hypothetical implications of offsetting views include forms of cessation, as well as other implications, that are arguably much worse than the implications of minimalist views (cf. 'Intense Bliss with Hellish Cessation' and 'Creating Hell to Please the Blissful').[231]

Are the cessation implications of experientialist consequentialist views a reason to reject *all* such views? One could argue that they are. Yet they may also be seen as a reason to *mind the gap* between consequentialist theory and practice. After all, experientialist consequentialists of every kind tend to justify strong norms against killing and violence *not* directly at the level of their preferred axiology, but at the level of practical decision procedures.[232]

fact that people report sincerely valuing things other than happiness and the absence of suffering, even when it is argued to them that such values could just be a conflation of intrinsic with instrumental value, often gives little pause to [such] utilitarians. But this is precisely the position a strongly suffering-focused utilitarian is in, relative to [offsetting] hedonists. That is, although this consideration is not decisive, [an offsetting] hedonist should not be convinced that suffering-focused views are untenable due to their immediate intuition or perception that happiness is valuable independent of relief of suffering. They would need to offer an argument for why happiness is indeed intrinsically valuable, despite the presence of similar debunking explanations for this inference as for non-hedonic goods [cf. 1.2].

[231] Knutsson, 2021b, sec. 3. For additional hypothetical implications of offsetting views, see, for instance, Vinding, 2020d, chap. 3.

[232] Cf. Mayerfeld, 1999, pp. 120–125. More on practical decision procedures below and in Chapter 2.

5.2.6 The Gap Between Theory and Practice

This section briefly highlights the gap between consequentialist theory and practice (before 5.3 does so at length). On the practical side, I argue that minimalist consequentialists, like other consequentialists, indeed should follow strong prohibitions against killing and violence in general.

The next subsections look at the following two questions, respectively:

1. Whether our practical anti-violence intuitions (strong and warranted as they are) might "miss their mark" in thought experiments that involve the cessation of causally isolated lives (be it 'minimalist cessation' or 'intense bliss with hellish cessation'), and whether this might constitute an additional bias in such thought experiments.

2. Whether there are experientialist minimalist reasons to strongly oppose even painless killing in practice.

5.2.6.1 Cessation and Our Practical Anti-Violence Intuitions

Regarding the first question, let us note from the outset that our practical intuitions are, of course, often correctly tracking the massive, negative effects that are associated with violence, killing, and dying in the real world.

At the same time, the hypothetical choice of 'minimalist cessation' involves *no* subjectively felt harm, *no* secondary effects for external beings, *no* loss of positive roles, and *no* uncertainty about the outcome whatsoever. This is highly unrealistic. Thus, it seems likely that our real-world adapted, anti-violence intuitions would (at least partially) miss their mark in this unrealistic thought experiment.

After all, it makes sense that our intuitions would treat death as a great bad in itself, even if we may on reflection think that its badness comes from these neighboring phenomena that are merely

often correlated with it. And even if we "abstract away" or intend to omit these phenomena in our hypotheticals, it seems likely that our intuitions on the badness of death are not easily moved by adding the magic words 'instant', 'all else being equal', or 'we are not affected by our choice in any way'.[233]

To be clear, many of these points are equally applicable to the offsetting case of 'intense bliss with hellish cessation'. Yet in that case, the cessation is preceded by an arbitrarily large hell (supposedly offset by the preceding bliss). So our harm-oriented intuitions may rightly raise concerns about the pre-cessation part of this offsetting implication, which does involve experienced harm.

5.2.6.2 Minimalist Reasons to Strongly Oppose Painless Killing

The hypothetical discussion above might spark the practical follow-up question: Without the concept of independent good, how can experientialist minimalist views oppose painless killing from a purely consequentialist perspective?

A common misconception surrounding minimalist views relates to their non-use of the concept of independent good. Why else would people take great pains to protect life and to even create new life? Yet we need not jump to the conclusion that these pains could be worthwhile *only* for the sake of some independent good. After all, minimalist views are perfectly compatible with the concepts of positive roles and positive lives, even if only in a relational sense, which can explain why we may rationally take great pains to protect and promote a variety of things even beyond their immediate preventive benefits.[234]

---

[233] 5.2.2; 5.2.4. Compare how people sometimes argue that our intuitions fail to closely track what is morally relevant in the case of the Repugnant Conclusion (3.3.2): "[The] unreliability of our intuition about the Repugnant Conclusion is due to a slight insensitivity in our intuitive grasp of the morally relevant factors." (Gustafsson, 2022.)

[234] More in Chapter 3 and Chapter 6.

The concept of positive roles also extends to positive norms, such as a general respect for autonomy and nonviolence, provided that these have overall greater preventive benefits than alternative norms do. So minimalist views can also imply a strong opposition to killing and violence for the sake of upholding positive roles, lives, and norms.[235]

Crucially, the upholding of overall positive norms will often justify the protection of overall negative lives (both on minimalist and offsetting views). Positive norms are not easily worth eroding for the sake of preventing more experiential bads at the individual level; on the contrary, they are among our most important resources to actively protect and develop.[236]

5.3 The Practical Side: Why We Should Not Seek to Create an Empty World

The hypothetical cessation response provided by minimalist views in 5.2.3 was based on assumptions that were completely unrealistic. Thus, we need to separately consider the altogether different practical question (**C**):

> Would minimalist consequentialism imply that it would be right to seek to turn our world into an empty one in practice, even by coercive means?

Why would anyone think so? A common route to such a conclusion may be what was previously called a "narrative misconception" of consequentialism.[237]

---

[235] Cf. 2.3.2.2; 6.3.

[236] See also 5.3.3 on the practical side.

[237] 3.5.1.

Let us next look at how such a misconception might arise, how it does more harm than good, and what a more accurate starting point for addressing the practical question would be.

5.3.1 Against Endstate-Oriented Utopianism

For minimalist views, an ideal world would be any world that is completely peaceful and free of problems, including an empty world. Thus, an empty world is one of the worlds that could be seen as a utopian outcome according to minimalist views. But does this observation have any practical implications? That is, should one be practically guided by the goal of creating an ideal world, or utopia, in the first place?

Utopian thinking does have its upsides. Empirical research supports the view that we may find psychological benefits from imagining our desired future world, such as a world free of painful experiences.

For instance, when people imagine their ideal society, they start to see more flaws in the current one and become more willing to help close the gap between the two.[238] This is a form of 'mental contrasting', which is a likely mechanism underlying the motivating effect of utopian thinking on social engagement.[239]

Yet utopianism can also be blind and dangerous, as exemplified by some of the greatest atrocities of the 20th century.[240] It may cause us to downplay or ignore the possibility of accidental harm: that we might make things worse, even when we think we are doing the right thing.[241]

---

[238] Fernando et al., 2018.

[239] Fernando et al., 2018.

[240] iep.utm.edu/totalita, "utopian".

[241] Cf. Wiblin & Todd, 2023.

5.3.1.1 Pitfalls of Utopianism

The pitfalls of utopianism outlined below are, in my view, some of the main reasons why people often draw fallacious practical conclusions based on consequentialist thought experiments, including when it comes to thought experiments about cessation in particular.

5.3.1.1.1 *"It Is the Only Way."*

Utopianism can take mental contrasting too far by focusing on an ideal image of how the future should go. However, the theoretically ideal future is not necessarily the best future to aim for in practice given the constraints and risks we are facing. For example, as argued below, it is plausible that the optimal goal to aim for in practice is not an ideal future but rather the prevention of worst-case outcomes.

5.3.1.1.2 *"We Must Get There."*

Consequentialism is not about seeking any particular "endstate" that we must realize in the future. Instead, the consequentialist understanding of "the end justifies the means" is that "the end" refers to the total consequences of our actions. It does not imply that reaching some ultimate destination would justify any means or risks necessary to get there.[242]

5.3.1.1.3 *"If Only Everyone Followed Along."*

Utopianism can be unrealistic due to requiring that everyone act in a certain way. Historically, all efforts to enforce utopian visions in a top-down manner appear to have failed.

When trying to bring about the best consequences, it is useful to distinguish between actions that are optimal from a "marginal realist" versus a "broad idealist" perspective.[243] Marginal realism is about how a relatively small group should optimally spend their

---

[242] Cf. the narrative misconception of consequentialism, 3.5.1.

[243] This distinction between "marginal realism" versus "broad idealism" comes from Vinding, 2022g, sec. 8.1.

limited resources in order to have the best impact on the margin.[244] The answer to that question *might*, but will not necessarily, approximate the answer to the "broad idealist" question, which is the question of how an entire society should ideally act if there were clear goal alignment between all stakeholders.

For instance, it is reasonable to think that the whole world should ideally shut down all factory farms. Yet what if a small minority of people were to attempt a violent shutdown of all factory farms? Most likely, the effort would fail. Additionally, it would carry a great risk of making things worse, such as by antagonizing this group and their cause in the eyes of more powerful groups.

At worst, taking unilateral, coercive action can lead to an overall more ruthless world, which would likely be much worse than "business-as-usual" in the long term.[245] A better strategy for small groups is arguably to seek broader and deeper support for their views, without coercive means, by adopting a cooperative approach.

5.3.1.1.4 "Seeking the 'Final Step' Over Absolute Expected Impact."

Utopianism can entail a kind of completionism or perfectionism, as if the "final step" toward utopia were especially significant.

By contrast, impartial consequentialism would not value the "final step" toward an ideal state of affairs any higher than any other beneficial step of equal magnitude. (This idea is baked into 'scalar utilitarianism', which understands rightness not in binary terms of "right" versus "wrong", but as a matter of degree.[246])

Related is the empirically documented effect called 'proportion dominance', which refers to people's common preference to help a higher *proportion* (that is, percentage) of individuals, even when the absolute impact remains the same. One study found that

---

[244] Cf. probablygood.org/core-concepts/marginal-impact.

[245] See Gloor, 2018, "Business as usual", and the related illustration.

[246] Cf. Norcross, 2006; Tomasik, 2015b.

people preferred helping 225 out of 300 lives rather than helping 230 out of 920 lives, which indicates proportion dominance even at the expense of absolute impact. Yet most people, when they reflected on it, agreed that one ought to prefer the higher absolute impact.[247]

Proportion dominance is highly relevant in our practical context. After all, the completionist vision of abolishing all suffering on Earth might intuitively override the more abstract and speculative aim of preventing suffering from spreading beyond Earth, even though the latter may involve far more suffering prevented in expectation.[248]

That is, when it comes to comparing "suffering on Earth" versus "(the risk of) suffering beyond Earth",[249] we are not just talking about ~300 versus ~900 lives, but about suffering on one planet versus suffering on a scale that is potentially multiplied by many orders of magnitude. This suggests that proportion dominance may represent a serious bias in our thinking about what to prioritize in practice to best reduce future suffering.[250]

5.3.1.2 A Better Alternative: Expected Value Thinking

If utopian thinking can diverge from impartial consequentialist thinking in at least the four ways outlined above, how could a consequentialist avoid those pitfalls without losing the benefits of utopian thinking?

Only 'naive' consequentialism recommends that we look at just the direct and immediate effects of our actions. By contrast, 'sophisticated' consequentialism recommends that we ideally estimate all the effects of our actions, including their indirect, long-term effects. This requires that we account for all the ways in

---

[247] Bartels, 2006.

[248] See Gloor, 2018, "Business as usual", and the related illustration.

[249] Cf. Tomasik, 2011.

[250] A similar point is made in Vinding, 2020d, pp. 144–145.

which forceful actions such as coercion or rule-breaking could cause more problems than they prevent.

A way to approximate this impossible ideal is to compare our potential actions in terms of their expected value, which here refers to the probability-weighted sum of the independently relevant consequences that they might have.[251]

When expected value is difficult to estimate directly (as it usually is), the way to make consequentialism feasible in practice may be to focus on indirect measures or heuristics that serve as useful proxies for what ultimately matters.[252]

A preliminary framework of proxies for reducing future suffering is presented in the book *Reasoned Politics* by Magnus Vinding. These proxies include greater levels of cooperation, increased and more impartial concern for suffering, and a greater capacity to achieve shared aims, including the reduction of suffering.[253]

Thus, we may estimate the expected value of various possible actions indirectly, by considering their likely effects on such proxy measures.

Overall, 'sophisticated' expected value thinking can, similar to utopianism, foster hope and collective action by highlighting ways in which things could go better (cf. mental contrasting). At the same time, it can help us sidestep the pitfalls of utopianism, as it constitutes a risk-sensitive way to account for a range of possible outcomes, including risks of astronomical suffering, and avoids blindly fixating on any single, vivid image of what the future should look like.

---

[251] Cf. Todd, 2021; probablygood.org/core-concepts/expected-value. Expected value thinking encourages us to ideally consider the full landscape of possible outcomes, including risks of making things worse. By contrast, endstate-oriented utopianism may neglect such risks, especially if the risks are not emotionally salient or easy to think about. See also Vinding, 2022m.

[252] Cf. multi-level consequentialism, 2.3.2.

[253] Vinding, 2022g, chap. 9, "Identifying Plausible Proxies".

5.3.2 Key Considerations for Estimating Practically Optimal Aims

This section looks at the key practical considerations for assessing whether minimalist consequentialism, combined with sophisticated expected value thinking, would recommend or discourage efforts to create an empty world.

5.3.2.1 Cooperation and Gains From Compromise

From a marginal realist perspective, minimalists have strong practical reasons to stay on cooperative terms with others.

As noted in the Negative Utilitarianism FAQ, attempting to realize an empty world would be a highly objectionable way to reduce suffering from the perspective of many other views. Additionally, from a minimalist perspective:

> The difference between "no future" (i.e. no Earth-originating intelligence expanding into space) and a decent future, where concern for suffering and thwarted preferences plays some role … is much smaller than the difference between a decent future and one that goes awfully wrong.[254]

Thus, minimalists have much more reason to steer the future away from going 'awfully wrong' and toward it going 'decently' — a goal that everyone can agree with — than to privilege an aim of 'no future'.

The various reasons to prevent conflict and antagonism are covered in brief and accessible ways in other sources.[255] One of these reasons is that when we accommodate each other's wishes

---

[254] Anonymous, 2015, sec. 3.2. See also Gloor, 2018, "Business as usual", with the related illustration.

[255] See Vinding, 2020e, as well as Tomasik, 2011, "Why we should remain cooperative". For a brief and accessible book chapter on the importance of cooperation for reducing suffering, see Vinding, 2020d, pp. 205–213.

and avoid costly fighting with each other, we can better achieve mutual gains from compromise, even if our altruistic goals may not always be fully aligned with each other.[256] That is, if people with different values can work together, this can enable everyone to better steer the future in desired directions.[257]

All else equal, basically everyone can agree that the reduction of suffering is an important aim. Given this agreement, a reasonable starting point is to work toward reducing suffering while standing on common ground between multiple value systems.[258] One way to do this is to focus on improving the expected quality of future lives conditional on their existence,[259] such as by reducing the risk of worst-case outcomes.[260]

5.3.2.2 Considerations Related to Wildlife, Evolution, and Space

I quote the following considerations essentially verbatim from Simon Knutsson,[261] and connect them with some related points made by others.

> 1: If merely all humans died, there would be room for more suffering wild animals (Tomasik, 2016), and humans would no longer be able to reduce wild-animal suffering, which we may do if we survive (Vinding, 2015).

These considerations make it at least unclear what the future of other beings on Earth would look like if all humans ceased to exist. On the other hand, human space colonization could also

---

[256] Tomasik, 2013b; Ord, 2015.

[257] Of course, there are also limits to how far one should go to "be nice to other value systems", depending on things like reciprocity, cf. Tomasik, 2014a.

[258] Baumann, 2020b.

[259] Tomasik, 2011; Vinding & Baumann, 2021, sec. 3.2.

[260] Tomasik, 2011; Baumann, 2017b.

[261] Knutsson, 2021b, sec. 4.

multiply the scale of wild-animal suffering.[262] Given such vastly higher stakes compared to Earth as we know it, possibly the best way to reduce wild-animal suffering (in expectation) is to convince the relevant future actors to not spread it in the first place.[263]

> 2: Even if all sentient beings on Earth died, beings that suffer could still evolve again on Earth (Acton and Watkins, 1963, 96; J. J. C. Smart, 1989, 44). Also, if humans survive, we may reduce suffering in other parts of the universe (Pearce, 1995, chap. 4, objection 32), or, at least, if we spread through space, it may result in less suffering than if other spacefaring civilizations do so instead (Tomasik, 2011).[264]

Suppose that our civilization would have both the motivation and the technical ability to prevent the possibility of life, and thereby all suffering, on our planet. From a minimalist point of view, would such "planetary euthanasia" be the practically optimal aim of such a powerful version of humanity?[265]

This is far from clear, because that very same civilization, given its high technical ability and value alignment, could also in some scenarios become a guardian against extreme suffering taking place elsewhere in the reachable universe.

Thus, the closer we are to being the kind of civilization that actually could prevent all suffering right here on Earth, the more we might also play more positive roles by ensuring that the vast resources of the reachable universe would not become fuel for generating astronomically greater suffering.

---

[262] Tomasik, 2014b.

[263] Tomasik, 2013a. A similar argument might also apply to factory farming (cf. Alene, 2022; Fai, 2022, sec. V).

[264] Cf. Vinding & Baumann, 2021, and grabbyaliens.com.

[265] Thanks to Nil for the question.

> 3: Similarly, if all humans or all sentient beings on Earth were killed, a new spacefaring civilization may eventually develop on Earth, and if it were to colonize space, it is an open question whether it would result in more suffering than if humanity were to expand into space (Tomasik, 2013c).

A key factor in the bigger picture is the difference in suffering that might result from one colonization wave versus another, including those that may stem from other civilizations.[266] From a suffering-focused perspective, if the resources of the reachable universe could be acquired by one or another colonization wave, then it is better (all else equal) that they be acquired by the one that would use them in the least suffering-conducive ways.

5.3.2.3 Risks of Astronomical Suffering

A reason to take seriously the considerations of cooperation, compromise, and empirical uncertainty is that they may have crucial relevance for what are called risks of astronomical future suffering,[267] also known as **suffering risks** or **s-risks**.[268] Roughly, s-risks can be understood as "events that would bring about suffering on an astronomical scale, vastly exceeding all suffering that has existed on Earth so far".[269]

From a minimalist perspective, a likely optimal goal to adopt in practice is to actively reduce s-risks.[270] This is because they are not extremely unlikely,[271] are bigger than present-day suffering in

---

[266] Vinding & Baumann, 2021.

[267] Tomasik, 2011.

[268] Baumann, 2017b. For a book-form introduction by the same author, see *Avoiding the Worst: How to Prevent a Moral Catastrophe* (2022).

[269] Baumann, 2017a. (A slightly different but practically similar definition is used in Althaus & Gloor, 2016; DiGiovanni, 2023.)

[270] Cf. Gloor, 2018, "Downside-focused views prioritize s-risk reduction over utopia creation"; Baumann, 2020a.

[271] Baumann, 2017b, "S-risks are not extremely unlikely".

expectation,[272] are neglected,[273] and there are ways we can reasonably reduce their probability and severity.[274]

In addition, the most promising interventions for s-risk reduction are often robustly beneficial for other aims as well.[275] For example, we can also better reduce near-term suffering and many other risks by achieving greater levels of cooperation, impartial moral concern, and capacity to steer the future in wiser directions.[276]

5.3.2.4 Strong Reasons to Prioritize Safer and More Widely Shared Aims

We have seen how minimalists have strong reasons not to prioritize an aim of 'no future'. For instance, doing so could greatly increase conflict and hostility between future actors, and indirectly increase risks of astronomical suffering.[277]

Instead, a better strategy is to prioritize safer and more widely shared aims, enabling greater cooperation between future actors, based on the significant common ground around the aim of preventing worst-case scenarios.

As argued in the Negative Utilitarianism FAQ:

> [Minimalists] will benefit more by cooperating and compromising with other value systems in trying to make the future safer in regard to (agreed-upon) worst-case scenarios, rather than by trying to prevent space colonization

---

[272] Baumann, 2017b, "S-risks outweigh present-day suffering in expectation".

[273] Baumann, 2017b, "S-risks are neglected".

[274] Baumann, 2017b, "How can we avert s-risks?"; Baumann, 2022.

[275] For some such interventions, consider Baumann, 2017b, "Broad interventions".

[276] Cf. the positive proxies for reducing future suffering discussed in 5.3.1.2; Vinding, 2022g, chap. 9.

[277] Tomasik, 2011, "Why we should remain cooperative"; Baumann, 2019, "Conflict and hostility".

from happening at all. It would be a tragedy if altruistically-concerned people split up into opposing factions due to them having different definitions of "doing what is good", while greed and bad incentives lead the non-altruistically-inclined people in the world to win the race. Instead, those who share *at least some significant concern* for the reduction of suffering should join together.[278]

5.3.3 A Safeguard Against Worst-Case Outcomes: Pragmatically Absolute Nonviolence

I'm as near as one comes to [being] a pacifist as is possible without being a pacifist. [Yes], there are exceptional circumstances in which violence may be unavoidable; we all know that life is messy. But other things being equal, I think the sanctity of life is a very good utilitarian principle because it promotes respect for other sentient beings. (David Pearce.[279])

Anti-harm ideas have inspired uniquely nonviolent practices for millennia. For example, many Jains and Buddhists aim to follow the principle of Ahimsa: never hurt another sentient being by word or deed.[280] Yet, impartial minimalism is not merely about minimizing our own personal "hurt-footprint". Instead, it recommends that we aim to minimize overall hurt for all sentient beings, regardless of the act-omission distinction.

Do minimalist views in practice diverge from absolute Ahimsa any more than do offsetting views? Other things being equal, all experientialist consequentialist views recommend that we cause the lesser hurt when it is the only way to prevent a greater

---

[278] Cf. Anonymous, 2015, sec. 3.2.

[279] Danaylov, 2013, 1:03:01.

[280] Cf. en.wikipedia.org/wiki/Ahimsa.

hurt. But compared to offsetting views, minimalist views are unique in saying that such situations are the *only* ones in which we could ever be justified in hurting others. Offsetting views provide more ways to justify hurting others (whether by act or omission), such as for the sake of creating independent goods that purportedly counterbalance the harm.[281]

When we zoom out from the personal "hurt-footprint" perspective, we may see minimalist consequentialist views as being based on a principle similar in spirit to that of Ahimsa, yet expressed in fully impartial terms: "The less sentient beings hurt, the better, regardless of the source."

At the same time, the abstract aim to minimize suffering is not by itself a directly suitable principle for guiding practical action.[282] In practice, to align with this aim, we need to follow more actionable principles, such as virtues and commonsense prohibitions that indirectly tend to bring about the best consequences.[283]

Additionally, in order to prevent our corruptible parts from opportunistically breaking such principles in self-serving ways, we likely need to internalize these principles deeply into who we are. (The road to hell is paved with good intentions that allow us to convince ourselves that we are in exceptional circumstances that warrant discarding the anti-harm precepts of commonsense morality. We rarely are.)

Finally, the sanctity of life (in roughly the way as understood in Ahimsa) is plausibly a key principle to protect as a safeguard against worst-case outcomes. This is in part because it promotes

---

[281] Cf. 5.2.5. It has also been argued that certain offsetting views are more likely, in practice, to lead to totalitarianism and to the use of coercive means than is a focus on reducing suffering. Specifically, Karl Popper, in his essay "Utopia and Violence" (1947/1986), defends such a claim in relation to views that justify present misery for the sake of realizing a future utopia that supposedly compensates for the misery. See also Popper, 1945, chap. 24; Danaher, 2018.

[282] Mayerfeld, 1999, pp. 120–125.

[283] Vinding, forthcoming, chap. 9, "Cultivating and Adhering to Virtues".

positive norms against aggression, and in favor of peace and co-operation.[284]

Overall, one of the best ways to counteract risks of future suffering may be to promote respect for all sentient beings in the form of unambiguously compassionate principles. These could be impartial forms of Ahimsa, promoting the path of least harm. They could also be called pragmatically absolute nonviolence and non-aggression.[285] Without such principles, the risks seem worse. With them, we have more hope.

---

[284] Vinding, 2020e. In many cases, nonviolent movements have also been more effective at achieving their altruistic aims; cf. Chenoweth & Stephan, 2011.

[285] More on pragmatic nonviolence and non-aggression in 6.3.

Part III

Positive Roles in the Real World

Chapter 6

Positive Roles of Life and Experience in Suffering-Focused Ethics

A common objection to suffering-focused ethics is that it contradicts our practically prioritizing other pursuits, such as seeking enjoyable and enriching experiences. To respond, this chapter aims to show that even if we assume a purely suffering-focused view, it would still be wise to recognize the highly positive and often necessary roles that other things may have for reducing suffering. Suffering-focused views may value these other things for different reasons, but not necessarily any less in practice, than do other consequentialist views. Moreover, in order to resolve tradeoffs between seemingly positive values, we may find great clarity in unpacking their causal relations to the suffering of sentient beings. A focus on reducing suffering can thereby help make different values commensurable with each other, and hence be a way to ground a framework for value prioritization.

6.1 Introduction

> *Suffering-focused ethics* is an umbrella term for moral views that place primary or particular importance on the prevention of suffering. Most views that fall into this category are *pluralistic* in that they hold that **other things** besides reducing suffering also matter morally.[286]

Most people agree that reducing intense suffering is important, other things being equal. When other things are not equal, we enter the realm of tradeoffs between values. And here it may seem as if suffering-focused views would override all other things at every opportunity to reduce suffering.

My aim in this chapter is to describe why this is not the case in practice, *even if* we assume a purely suffering-focused view that is ultimately concerned only with reducing suffering.

In this chapter:

- I assume a monist view, in which only suffering has *independent* value (that is, 'intrinsic disvalue'), and where other things can have positive value only by virtue of their tendency to reduce this negative value.

- I use the term 'other things' to mean values such as:

    ○ Autonomy (6.3.1); maintaining stable ecosystems (6.3.2); cultural diversity (6.3.3).

    ○ Wellbeing and flourishing (6.4.1); overcoming challenges (6.4.2).

    ○ Exploration of helpful outlooks (6.4.3); growth and learning (6.4.4).

    ○ Depth and variety of experience (6.4.5); social relations (6.4.6).

---

[286] Gloor, 2016, bold emphasis mine.

○   Meaning and positive narratives (6.4.7).

In addition to the values mentioned above, there are also con-sequentialist defenses of other popular intrinsic values, such as rights and virtues, yet these are beyond the scope of this chapter.[287]

The next section will briefly describe how we might better in-tuitively appreciate the positive roles discussed in this chapter, namely by avoiding only seeing them through the potentially nar-row lens of 'instrumental value'.

6.2 Possible Misconceptions About Instrumental Value

Below are some ways in which it might be counterproductive to think of various valuable things as "merely instrumental" in ev-eryday life:

- **Misleading connotations of being used as a mere tool**. The term 'instrumental value' may increase the likelihood that we would perceive the value of something or some-one as "merely instrumental… merely a tool", or as if they would have positive value and moral worth "only" by virtue of being "used" to serve the goals of others. These connotations are misleading in the context of impartial suffering-focused ethics, in which:

  ○   All beings are taken into primary consideration based on their capacity to suffer.

  ○   Beings and things alike can play positive roles for reducing suffering without anyone "using" them this way. Moreover, many beings and things can play highly positive roles for reducing suffering

---

[287] For sources that present suffering-focused defenses of rights, see Vinding, 2022g, pp. 79–80. For a suffering-focused defense of virtues, see Vinding, forthcoming, chap. 9, "Cultivating and Adhering to Virtues".

without anyone realizing it, and do so in ways that are a lot more systemic and far-reaching than the "direct-use" utility of what we often call a tool or instrument in everyday life.

○ Impartial instrumental value is ultimately about the overall benefit of all beings, not about serving some restricted subset of beings for whom everyone else should serve as a tool (cf. the misleading connotation of being used purely for the goals of others).

● **A second-rate, lesser kind of value**. If we perceive a thing as "merely" instrumental, it may feel as if its value could at any time be overridden by a more important kind of value, namely by independent or intrinsic value. Yet this need not be true, since instrumental value is ultimately measured on the same scale as independent value or disvalue. After all, the positive roles of a thing are often greater than its independent moral weight.

○ For instance, sometimes learning experiences can entail great pains. Yet these learning pains are often dwarfed by their highly worthwhile positive roles, such as how they can help us grow into effective reducers of even greater pains.

● **Falsely thinking that we can readily grasp the full magnitude of its value**. Appreciating the overall instrumental value of a simple tool like a hammer can already be a difficult task. It is much more difficult still to grasp the value of things that play positive roles in more complex, indirect, and systemic ways, such as the norms of nonviolence and a predictable respect for each other's autonomy.

○ Even a theoretically monistic consequentialism implies that it is practically best to treat many of our culture's widely held intrinsic values as valid

moral heuristic*s* to follow and respect — at least until they run into edge cases or conflicts with each other, at which point we may want to carefully unpack their roles under a common standard of value.

6.2.1 Better Alternatives?

If we can avoid problems such as those above, it seems useful to continue using the term 'instrumental value'. Alternatively, some of these problems may be easier to avoid if we instead think of this value in some other terms, such as extrinsic value, relational value, or positive roles. As an experiment in framing, I use the terms 'relations' and 'roles'.

6.3 Life and Diversity

Suffering-focused views may appear as if they would ignore or override values widely held to have almost sacred status, such as a general respect for personal autonomy, life, and ecosystems. However, a careful account of how to best reduce suffering in practice suggests that it would be self-defeating to seek to help others via means that go directly against such deeply and widely held principles and values, many of which are also quite aligned with reducing suffering to begin with.

6.3.1 Autonomy

Most of us have a strong need for making independent decisions: we want choice and predictability in our lives instead of being restricted or maneuvered by others. We easily perceive forceful limitations on our autonomy as being offensive or manipulative, which may be based on historically justified skepticism of anyone being able to handle our personal affairs better than we ourselves.

Moreover, an unmet need for autonomy can by itself cause great suffering.[288]

In terms of motivating others to reduce suffering, probably the most effective strategy is to appeal to people's own free choice rather than to use more forceful methods, whether they be forceful rhetoric or actual force. Beyond being less effective, such forceful methods also come with a greater risk of causing a harmful backlash.

When allowed to act freely, most of us are already interested in avoiding intense suffering for ourselves, and many will also make efforts to help or at least not harm others. Of course, sometimes we may exercise our autonomy in harmful ways. Yet this alone does not justify strict limitations on freedoms of speech, movement, or self-direction, as such limitations may cause far more suffering all things considered. For example, when the powerful allegedly "know better", and power becomes corrupted, the result tends to be suffering for the masses.

Skepticism of top-down control may be the main reason to respect autonomy on any view, including suffering-focused views. Top-down control is often abused or impractical even if well-intentioned, which has led to the separation of powers within governments as a safeguard against power becoming concentrated in harmful ways.

A general degree of respect for individual autonomy may thus reduce suffering by upholding at least two important freedoms: (1) the freedom to protect our own interests and life plans from external mismanagement, and (2) the freedom to organize and speak up against perceived harmful developments, practices, or corruptions in society at large.

Because exceptions to these freedoms have historically been all too easy to abuse, it is arguably best to maintain a high and consistent standard of autonomy for all beings capable of informed choice. To the extent that autonomy is used to harm others,

---

[288] On the strength and universality of a need for autonomy, see Yu et al., 2018.

we can focus preventative measures on specific harmful actions instead of limiting autonomy as a whole. Many societies already address the downsides of high autonomy on such a case-by-case basis, without losing the upsides of high autonomy for preventing suffering.

Suffering-focused views…

- ***… can justify*** many of the commonly accepted limits to autonomy. These may include cases of reduced capacity for informed choice (medical ethics)[289] or cases of protecting the public by preventing someone from severely harming others (criminal law)[290].

- ***… support*** giving people the autonomy to contribute to the reduction of suffering as they best see fit, including developing their unique skills and gravitating to the roles that they are best suited for.

- ***… can allow*** us the freedom to make mistakes, because this is often the lesser harm compared to the alternatives.

6.3.2 Ecosystems

Purely suffering-focused views can sound as if they are against life in principle. But practically speaking, their implications are more complex than that. To guide suffering-reducing action, we need to carefully account for the reality that we find ourselves in.[291]

Rather than ask whether we would wish that suffering beings had never evolved, it is more useful to consider how we can best help them from our current situation. And the current world can be complex, containing factors such as multiple interacting value systems and their dependence on stable ecosystems for most of

---

[289] Cf. en.wikipedia.org/wiki/Medical_ethics.

[290] Cf. en.wikipedia.org/wiki/Criminal_law.

[291] Cf. Chapter 5.

their long-term goals, but also a lot of neglected wild animal suffering within those ecosystems.[292]

There are no simple answers to the question of how to best reduce wild animal suffering. Its scale and tractability depend on various moving parts, such as what values and political and economic forces will end up shaping the future. In a world where people hold diverging values, there are strong reasons to cooperate with others, because mutual conflict would probably lead to worse outcomes overall.[293] For example, conflict might nudge the long-term future into a more ruthless and competitive direction, hindering the progress of most long-term goals, including suffering-focused goals.[294]

Cooperating with others might be especially relevant when it comes to how we can best approach the problem of wild animal suffering. Arguably, a top priority for suffering-focused views is to promote peace and compromise so as to reduce the risks of astronomical suffering (s-risks) that could result from the worst forms of our civilization in the long term.[295] Thus, to the extent that stable ecosystems are necessary for a peaceful civilization, suffering-focused views would ideally seek agreement with others on how to best reduce the suffering of individual beings within existing ecosystems.[296] After all, a great many if not most moral views already imply that non-human animals matter morally, and that their suffering deserves serious consideration.[297]

---

[292] en.wikipedia.org/wiki/Wild_animal_suffering.

[293] Cf. Tomasik, 2013b, 2014a; Vinding, 2020e.

[294] Cf. Baumann, 2020b.

[295] Baumann, 2017b. For a book-form introduction by the same author, see *Avoiding the Worst: How to Prevent a Moral Catastrophe* (2022).

[296] Cf. Tomasik, 2013a, "Global stability".

[297] Animal Ethics, 2020.

Suffering-focused views…

- *… **question*** the romanticized, sanitized view of nature as good and harmonious, especially since most of the animals living there endure fates much worse than what people would find acceptable for their pets or children.

- *… **raise concern about*** even a small chance of spreading wild animal suffering onto other planets, as in certain futuristic terraforming scenarios.[298] This also applies to the creation of artificial suffering in the future, as well as to factory farming.

- *… **benefit from cooperation*** among people who hold different values, since avoiding intense suffering is already a shared goal for most people, and maintaining peace may be a uniquely promising strategy for reducing risks of astronomical suffering in the long term.

6.3.3 Cultural Diversity

Suffering-focused views may seem to be in tension with respect for cultural diversity. However, there are many good reasons for suffering-reducers to approach the world's cultural diversity with curiosity and respect, including epistemic modesty as well as the recognition that many traditions may have developed time-tested ways to mitigate and relate to intense suffering that we may not know about.

For example, living traditions such as Buddhism and Jainism predate modern medicine by millennia in investigating how to actively reduce suffering, both for oneself and others, and science has only recently started to study the psychological and pro-social benefits of compassion and non-attachment.[299]

---

[298] Tomasik, 2014b.

[299] Seppälä et al., 2017; Whitehead et al., 2018.

Cultural traditions may thus contain neglected wisdom for reducing suffering, and plausibly the best way to create helpful innovations in general is to combine the ideas found across diverse minds and cultures.

Moreover, any ethical view that aspires to universality, including impartial suffering-focused views, can value cultural diversity as a safeguard against the emergence of a harmful intellectual monoculture.[300] Diverse cultures can provide ideas and critiques that may be difficult to perceive from within a narrow cluster of views, due to each having their own biases and blindspots, which is all the more reason to engage in healthy dialogue with those who have different outlooks.

Suffering-focused views…

- *… can value* different cultural practices based on the degree to which they might help prevent suffering, but would still be critical of harmful practices such as those that involve ritualized violence.

- *… can support* a general respect for other cultures not only to maintain peace and cooperation, but also for the implicit adaptations of different cultures for reducing suffering that may still be poorly understood.

- *… may need* the perspectives of various cultures for optimally reducing suffering across different contexts. For example, experimental psychology is still mostly based on participants from "WEIRD" cultures, casting doubt on whether its models will generalize to most of the world.[301] And probably no single culture contains all of humanity's helpful insights for reducing suffering, which suggests a need to integrate insights between cultures.

---

[300] On monoculture as a possible problem within the effective altruism community, cf. Kuhn, 2013, "Monoculture".

[301] Henrich, Heine, & Norenzayan, 2010a. The acronym stands for Western, Educated, Industrialized, Rich, and Democratic.

6.4 Valuable Experiences

An objection sometimes raised against suffering-focused views is that they seem to leave too little room for the valuable experiences in life, contradicting our everyday experience of what makes life worth living. Perhaps such a "negative focus" feels like a distraction from, or even a threat to, what one may see as more centrally important, such as growth, connection, or a sense of positive meaning. But suffering-focused views can imply that those things are important to prioritize for their positive roles for most effectively helping others.

6.4.1 Wellbeing as a Resource

When we are under a lot of stress, it is easy to lose our balance and fall into a downward spiral. At the bottom, it can take years to restore our supportive daily habits and capacity for effective work. From a lifetime productivity perspective, severe burnout is very much worth avoiding, even at the seemingly high cost of emptying our calendar of all the short-term work that we maybe could squeeze in, but not in sustainable ways. This is particularly true when we are still young, as peak productivity tends to come later in life.[302]

The better our work-rest balance is on a daily basis, and especially in the long term, the more we can afford to work. And to work from day to day at all, we need to maintain a level of wellbeing that provides a safety margin against burning out. In other words, we want to continuously create distance from a downward spiral. Any uplifting or supportive experience can be a great way to create this distance.[303]

Thinking about suffering can be distressing. Most of us would rather fill our days with something more pleasant than imagining

---

[302] Todd, 2023.

[303] See also Leighton, 2017, "wellbeing", "renew".

worst-case scenarios and how to avoid them. But suffering-reducers need the resilience to face these questions head-on. Many of our enjoyable experiences are effective ways to get our minds off these problems and replenish our ability to get back at solving them. Even if we do not consciously "use" these experiences as "mere tools" in this way, they can have these benefits anyway.

By frequently enjoying what we find replenishing, we can gather the skills, knowledge, and life experience to understand and apply the optimal effort that would best help others. And by maintaining a high level of personal wellbeing, we can sustainably apply ourselves to solve neglected problems for decades, without burning up our capacity to do so.[304]

Suffering-focused views…

- *… value* personal wellbeing as an indicator of low personal suffering, and they can prioritize the creation of wellbeing as a *way toward* greater help in the long run, but purely suffering-focused views would not say that anyone's wellbeing could by itself offset or counterbalance intense suffering elsewhere.

- *… are not about* maximizing personal wellbeing or distance from burnout, but rather about optimizing these things for overall helpfulness. Contrary to motivated reasoning by our short-term pleasure-seeking parts, a life of optimal helpfulness may not always require high levels of momentary pleasure or excitement (though it can still have room for these things). Just like with money or any other resource, there is a point where our personal wellbeing is no longer the main limiting factor to what we can do, and so we can afford to help others while keeping a safe distance from burning out. This may also be a more

---

[304] For additional resources on sustainable activism, see Vinding, 2017; forthcoming, "Reducing Extreme Suffering in Healthy Ways".

reliable path to personal life satisfaction than to always focus on improving our personal wellbeing.

- *... **are not about*** avoiding all personal suffering, but tolerating that which is worthwhile in order to best help reduce overall suffering. Sometimes the optimal path of maximal "net helpfulness" may contain great difficulties that require a large reserve of wellbeing to get through. Wellbeing is thus a key resource to focus on, but we also want to look out for opportunities to invest it in once we ourselves are in a healthy place from which we can help others.

6.4.2 Moving in the Right Direction

Most of us have no choice over our childhood environment or upbringing. By the time we develop a sense of agency and start thinking about global issues, it can feel overwhelming to realize that our deeply ingrained habits and lifestyles are indirectly causing a lot of suffering. Combined with an already busy life full of personal challenges, it can feel more painful than empowering to start reducing our reliance on habits and production chains that may be net harmful. Only the most fortunate get to choose their jobs and limit their consumption so as to minimize their suffering footprint for others.

However, there is positive value in any journey of overcoming our dependence on others' suffering and toward becoming increasingly net helpful. Bit by bit, we can increase our degree of freedom and victory over harmful dependencies. With careful research, we may also identify amazing opportunities to reduce suffering that are available to us.

The more hopeless the situation, the more inspiring the example of turning it around or pushing through, even partially, for all the others who are facing similar challenges. Even if we do not make it all the way, we can still share our story, and others can continue with more guidance than we had, with a better view of the path going forward.

Suffering-focused views…

- *… do not mean* that we should give up on our lives if they cause or contain a lot of suffering. In many cases, these are precisely the situations that countless others are also struggling with. There is value in first-hand exploration of the problems and in sharing even small wins that others may not know are possible.

- *… do not imply* that we should bear anything in order to create inspiring survivor stories of overcoming and victory for others. After all, the "bottom line" is to reduce suffering instead of always clinging to dwindling hope that a cure for chronic and severe conditions will be just around the corner.[305]

- *… are context-sensitive* regarding the options that may or may not be available in our life situation. There are no absolute demands of what we must achieve regardless of our health, wealth, or the environment we find ourselves in. But we can always do our best to steer the future into a direction of less suffering with the tools and options that we have. Everyone climbs their own way up the mountain, and even a partial path over difficult terrain can inspire others to find an easier way up from similar starting points.

6.4.3 Exploring Helpful Outlooks

We humans have the unique ability to actively develop and experiment with different ways of looking at suffering. Over time, many people have made good use of this ability, sometimes passing the resulting practices and outlooks on to future generations. Some time-tested outlooks can be found under traditions and philosophies such as Buddhism and Stoicism. Even many of our modern

---

[305] For more on euthanasia from an impartial suffering-focused perspective, see, for instance, Vinding, 2022g, p. 205.

psychological tools for relating to suffering can be traced back to such traditions, and we may yet combine parts of them into still more useful or teachable outlooks than what are currently available.[306]

For the overall project of reducing suffering, it is worthwhile to spend time on finding the kind of open mindset that enables us to acknowledge our own and others' suffering, and which best helps us reduce it over all time. For example, we want to build an outlook that helps us be mindful of the big picture instead of over-reacting to whatever may seem like the totality in the here-and-now. With a large perspective, even apparent defeats and failures can be seen, not only in hindsight, but already in the present, as opportunities for growth whose time has come. As some Stoics say, the obstacle can become the way.

Suffering-focused views…

- **… *value*** trying out different mindsets and developing the equanimity to face our difficulties, especially if mental resources might be a key bottleneck for our ability to help ourselves and others.

- **… *may recommend*** humor, spontaneity, and flexibility in many situations as fitting ways to meet life's challenges.[307]

6.4.4 Safe Ways to Learn

Even if we ultimately seek to reduce intense suffering for all beings, this does not require that we maintain a myopic focus and prioritize only those activities that would help others directly. A single-minded or Spartan life could leave too little room for the ways in which we spontaneously learn the most (and most effort-

---

[306] Shapiro & Weisbaum, 2020; Cavanna et al., 2023.

[307] For more on humor as a virtue in the context of reducing suffering, see Vinding, forthcoming, sec. 9.11, "Humor and Optimism".

lessly), such as through flow-states guided by feelings of immediate fulfillment. It might also leave too little room for activities and learning experiences that are only indirectly related to the goal of reducing suffering, even if they are in fact crucial in the bigger picture. After all, many activities can prepare us for challenging future tasks without containing any obvious reference to such tasks. Most of us learn social skills this way, never picturing what they might be useful for in ten years' time. Many fundamental skills can similarly be learned through unguided exploration, pretend play, or games. To constantly hold our ultimate aim in mind will often distract us from optimally learning the skills on offer, including skills that may be necessary on the path of least suffering.[308]

Of course, at some point we need to directly learn about the reality of suffering and how we might best reduce it. Yet we might only need a brief sample of how bad it can get. After first learning that fire burns, we may coordinate our movements around it for the rest of our lives.

For high-level tasks that affect the lives of others, it is best to develop the relevant skills first in "safe mode" before we move on to more serious and consequential settings. High-stakes themes, such as philosophical, political, and social issues, are often best explored first in "low-stakes" sandbox environments, where we can get disproportionate learning value with few backfire risks. For example, while casual discussions or debates with friends may at first glance seem suboptimal from an impact-focused perspective, they may also be helpful stepping stones before we enter the high-stakes arenas, such as public discourse, where taking an actual stance can have broader consequences, for better and worse.

---

[308] Cf. the book *Why Greatness Cannot Be Planned: The Myth of the Objective* by Stanley & Lehman, 2015.

Suffering-focused views…

- *… can value* open-ended exploration (to a degree), because learning is often most efficient when guided by feelings of intrinsic motivation and curiosity.

- *… would not value* endless learning or recreation purely for its own sake. After all, at some point, it becomes worthwhile to focus our attention on the high-stakes issues, and to prioritize actively finding ways in which we can best apply our knowledge and skills to nudge the world in a better direction.

6.4.5 Understanding Others

To navigate effectively in the world, we may need to increase not only our knowledge and technical know-how, but also our own range of experiences and perspectives. Otherwise, we may in many cases fail to empathize with the central needs and aspirations of others.

For example, it is valuable for suffering-reducers to understand the most universal human experiences driving much of our behavior. These include the ups and downs of social status; the dynamics of attachment and bonding; and our needs for autonomy, competence, and relatedness.

Psychological and social dynamics may be the main factors that determine which messages and strategies for reducing suffering will be well-received and politically realistic to implement. Thus, it is valuable for agents of change to understand the needs and incentives that determine which practical measures for reducing suffering can be adopted in the real world.

Suffering-focused views…

- *… value* psychological exploration and empathy, which generally grow our capacity to understand and help both ourselves and others.

- *... **may be surprisingly supportive of** fictional works, such as realistic or thought-provoking novels, as well as nonfiction works about key social and historical topics. After all, such sources may be safe ways to learn about real-world social patterns without getting us entangled in actual drama.*[309]

6.4.6 Community and Connectedness

A lot of common and prolonged suffering could be avoided if people had close friendships, or at least didn't feel so alienated or frustrated with their social environment. Moreover, a big part of the willingness to even consider the needs of others in the first place may come from a social norm to reciprocate or pay forward the support that we have experienced from others.

If people have strong unmet needs of their own, they may not find it emotionally appealing to serve a universal cause such as the reduction of suffering. At worst, they might actively cause more suffering, perhaps out of frustration or a lack of meaningful experiences with others.

Thus, a valuable part of reducing suffering is to foster the preventative and empowering effects of increasing community and connectedness in society. This could help more people feel like they're at home in our world, and that they can afford to care about the needs of others.

Another reason why a sense of community is important is that it can help facilitate complex coordination. For example, to identify the most promising interventions for reducing suffering, we will need to combine insights from a wide variety of fields. To make this happen, we need to create and maintain an interdisciplinary network that is willing to spend many decades learning how to best reduce suffering. And a sense of community is likely vital for holding such multi-decade efforts together.

Suffering-focused views...

---

[309] Cf. Mumper & Gerrig, 2017; Dodell-Feder & Tamir, 2018.

- ... *value* the beneficial effects that can stem from experiences of community and connectedness, both individually and in society at large. Experiences of connection may be a precondition for our wellbeing, enabling us to care about and effectively serve the needs of others.

- ... *highlight* the importance of preventing complex social problems from becoming entrenched while we still can. A prime example of this may be social dissatisfaction, as deeply ingrained feelings of anger or loneliness can lead to self-reinforcing patterns where we blame everything on others and discount each other's suffering. At worst, the suffering of others may come to be seen as desirable, even if there once was a desire for connection.

- ... *may recommend* creating and inviting people into communities focused on reducing suffering, embodying sentiments of a collective project, such as: "Let's work together to reduce suffering."[310]

6.4.7 Meaning and Motivation

Many paths can lead one to suffering-focused ethics. Only some of the paths are based on a sense of duty or obligation to prevent the worst experiences. Other paths can arrive at the same goal of reducing suffering through a more positive or hopeful sense of what our lives are worth living for, even in the face of extreme adversity.

One common positive motivator is a sense of expanding care or compassion for all sentient beings. Many of us feel compassion and an urge of caretaking toward ourselves and loved ones. The same tendency of caretaking can shrink or grow, based on our capacity to consider the needs of others together with our own. At the limit, we may see alignment with suffering-focused ethics as a way of living for compassion: a life aimed at helping sentient

---

[310] As expressed in a tweet by psychologist Paul T. P. Wong, twitter.com/PaulTPWong/status/1392966664803831814.

beings as much as we can, and feeling as satisfied as possible with how we spent our time.

Another positive motivation is to live for a maximally just, fair, or beautiful world. We may recognize that any instance of extreme, involuntary suffering, no matter when, where, and by whom it is experienced, means that we are nowhere close to such a heavenly world. We may feel that the best way to work for the ideal of a "heaven on earth" is to reduce hell, by sparing as many beings as possible from having to experience the unbearable.

Suffering-focused views…

- ***… are not necessarily*** "all negative", even if suffering it-self mostly is. By focusing on experiences of care, com-passion, justice, ease, or lightness, we can bring together even those parts of ourselves that are motivated by a posi-tive vision of what we're living *for*.

- ***… can acknowledge*** that suffering itself can also have up-sides; after all, our personal suffering can give us great clarity, direction, and motivation for how to be useful to others.

- ***… can give meaning*** to our lives, especially given that we are in a rare and privileged position from which we can help beings who are unable to help themselves. At mo-ments of despair, we may feel the root problem of invol-untary suffering common to all sentient beings, and push ahead for the benefit of all beings. To conclude, even secular helpers may find meaning in 8th-century Buddhist philosopher Shantideva's words: "As long as space en-dures, as long as sentient beings remain, until then, may I too remain and dispel the miseries of the world."[311]

---

[311] Gyatso, 2002, p. 125. For more on Shantideva, see Goodman, 2016. For more on the alignment between meaning, motivation, and reducing suffering, see *Compassionate Purpose: Personal Inspiration for a Better World* (Vinding, forthcoming).

Chapter Summaries

Below is a brief summary of the main points of each chapter in the book.

1. "Minimalist Views of Wellbeing":

   ➢ Personal wellbeing is often defined as "the balance of that which is good for oneself over that which is bad for oneself".

   ➢ We may be skeptical of such 'good minus bad' views of wellbeing due to the many reasons to doubt the offsetting premise — that is, the premise that independent bads can always be counterbalanced or offset by a sufficient addition of independent goods.

   ➢ This premise is rejected by all minimalist views of wellbeing. These include experientialist views, where wellbeing is the degree to which we are free from experiential sources of illbeing (such as suffering, disturbance, or a visceral non-acceptance of our current experience), as well as extra-experientialist views, where wellbeing is also affected by preference frustrations, interest violations, or objective conditions.

2. "Varieties of Minimalist Moral Views: Against Absurd Acts":

➢ Minimalist moral views are sometimes alleged — at least in their purely consequentialist versions — to recommend absurd acts in practice, such as murdering individuals or choosing not to save people's lives so as to prevent their future suffering. Yet there are various reasons why the most plausible versions of minimalist moral views — including their purely consequentialist versions — do not recommend such acts.

➢ These acts would be opposed by minimalist versions of nonconsequentialist views, such as virtue ethics, deontology, social contract theory, and care ethics.

➢ Consequentialist reasons against such acts can be derived from extra-welfarist and extra-experientialist axiologies, which may consider violence or violation to be inherently bad, as well as from rule and multi-level consequentialism, which highlight the instrumental reasons for respecting autonomy, cooperation, and nonviolence.

3. "Minimalist Axiologies and Positive Lives":

➢ Minimalist axiologies define betterness in entirely relational or 'instrumental' terms, namely in terms of the minimization of bads such as suffering.

➢ These views avoid many problems in population ethics, while retaining a plausible notion of *relational* positive value. Yet the minimalist notion of (relationally) positive value is entirely excluded by the standard, restrictive assumption of treating lives as isolated value-containers.

➢ Minimalist views become more intuitive when we adopt a relational view of the overall value of individual lives, that is, when we don't track only the causally isolated "contents" of these lives, but also their (often far more significant) causal roles.

4. "Minimalist Extended Very Repugnant Conclusions Are the Least Repugnant":

> ➢ It has been argued that certain "repugnant conclusions" are an inevitable feature of any plausible axiology.

> ➢ Yet at least some minimalist views avoid these repugnant conclusions. Moreover, based on a 'side-by-side' comparison of different views, a strong case can be made that offsetting views share all the most "repugnant" features of minimalist views while introducing additional sources of repugnance.

> ➢ The comparison suggests that the conclusions faced by minimalist views are the least repugnant and the most plausible overall.

5. "Peacefulness, Nonviolence, and Experientialist Minimalism":

> ➢ For purely experience-focused and consequentialist versions of minimalist views, an ideal world would be any perfectly peaceful world, including an empty world.

> ➢ When it comes to theoretical implications about the cessation and replacement of worlds, one can reasonably argue that offsetting views have worse implications than do minimalist views.

> ➢ Zooming out from unrealistic thought experiments, it's crucial to be mindful of the gap between theory and practice, of the pitfalls of misconceived consequentialism, and of how minimalist consequentialists have strong practical reasons to pursue a nonviolent approach and to cooperate with people who hold different values.

6. "Positive Roles of Life and Experience in Suffering-Focused Ethics":

> ➢ Even if we assume a purely suffering-focused view, it's wise to recognize the highly positive and often necessary roles that various other things may have for the overall goal of reducing suffering.

> ➢ These include the positive roles of autonomy, cooperation, and nonviolence, as well as our personal wellbeing and valuable experiences.

> ➢ Suffering-focused moral views may value these things for different reasons, but not necessarily any less in practice, than do other moral views.

Recommended Resources

Books

➢ *Avoiding the Worst: How to Prevent a Moral Catastrophe* (2022) by Tobias Baumann

➢ *The Tango of Ethics: Intuition, Rationality and the Prevention of Suffering* (2023) by Jonathan Leighton

➢ *Suffering and Moral Responsibility* (1999) by Jamie Mayerfeld

➢ *Suffering-Focused Ethics: Defense and Implications* (2020) by Magnus Vinding

➢ *Reasoned Politics* (2022) by Magnus Vinding

➢ *Essays on Suffering-Focused Ethics* (2022) by Magnus Vinding

➢ *Compassionate Purpose: Personal Inspiration for a Better World* (forthcoming) by Magnus Vinding

Essays and Articles

Magnus Vinding's blog:

➤ magnusvinding.com

Essays on Reducing Suffering by Brian Tomasik:

➤ reducing-suffering.org

Simon Knutsson's writings on ethics:

➤ simonknutsson.com

Organizations

Center for Reducing Suffering (CRS)

➤ centerforreducingsuffering.org

Center on Long-Term Risk (CLR)

➤ longtermrisk.org

Organisation for the Prevention of Intense Suffering (OPIS)

➤ preventsuffering.org

Acknowledgments

Many people have played supportive and formative roles in the creation of this book. First of all, I am deeply grateful to David Pearce for his many decades of advancing impartial compassion for all sentient beings. I still remember the inspiring impact it had on me when I first encountered his works online in 2011, which I believe planted the first seeds of my and many others' decision to devote our careers to reducing suffering.

This book was made possible thanks to the supportive work of my colleagues at the Center for Reducing Suffering (CRS) and the donors who enable it all in the first place. I am also grateful to the Center on Long-Term Risk (CLR) for a networking event where I could meet for the first time many people from the broader suffering-focused community.

For their feedback on earlier drafts, I thank Riikka Ajantaival, Tobias Baumann, Timothy Chan, Eleos Arete Citrini, Anthony DiGiovanni, James Faville, Aaron Gertler, Lukas Gloor, Barry Grimes, Simon Knutsson, Jonathan Leighton, Winston Oswald-Drummond, David Pearce, Michael St. Jules, Roger Thisdell, Brian Tomasik, Emma Tulanova, and anonymous commenters.

I am also grateful to David Althaus and Kaj Sotala for helpful conversations, and to Clark Wolf and Jonathan Leighton for sharing their thoughts on the final draft.

This book combines many ideas that have previously been published and discussed in detail by others. I would like to thank

all those people, and I hope that my notes and references throughout the book highlight the extent to which I am drawing on a large body of existing work.

I am most deeply grateful to my CRS colleague and editor, Magnus Vinding, who has provided detailed feedback and guidance throughout the process. His books and writings have had a major influence on my thinking, and I feel lucky to have had him as my editor. Thus, by extension, I wish to thank all the people who enable Magnus to do what he does.

Finally, I am deeply grateful to my wife Riikka for her love and support for the past decade. I love our life together and I'm curious what will happen next.

References

Acton, H. B. & Watkins, J. W. (1963). Symposium: Negative utilitarianism. *Proceedings of the Aristotelian Society, Supplementary Volumes, 37*, 83–114.
doi.org/10.1093/aristoteliansupp/37.1.83

Alene. (2022). Who is protecting animals in the long-term future?
forum.effectivealtruism.org/posts/JSEencwkDuwWzYny8

Alexander, L. & Moore, M. (2021). Deontological ethics. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Winter 2021 Edition). Stanford University.
plato.stanford.edu/archives/win2021/entries/ethics-deontological

Althaus, D. & Gloor, L. (2016). Reducing Risks of Astronomical Suffering: A Neglected Priority.
longtermrisk.org/reducing-risks-of-astronomical-suffering-a-neglected-priority

Animal Ethics. (2020, July 14). *Ethical theories and nonhuman animals* [Video].
youtube.com/watch?v=WsXzdi7zSOo

Anonymous. (2015). Negative Utilitarianism FAQ.
utilitarianism.com/nu/nufaq.html

Armstrong, S. (2019). The Very Repugnant Conclusion.
lesswrong.com/posts/6vYDsoxwGQraeCJs6

Arrhenius, G. (2000a). An impossibility theorem for welfarist axiologies. *Economics & Philosophy*, *16*(2), 247–266.
doi.org/10.1017/S0266267100000249

Arrhenius, G. (2000b). *Future generations: A challenge for moral theory* (Doctoral dissertation). Uppsala University.
diva-portal.org/smash/record.jsf?pid=diva2%3A170236

Arrhenius, G. (2003). The Very Repugnant Conclusion. In K. Segerberg & R. Sliwinski (Eds.), *Logic, Law, Morality: Thirteen Essays in Practical Philosophy in Honour of Lennart Åqvist* (pp. 167–180). Uppsala University.
philpapers.org/rec/ARRTVR

Arrhenius, G., Ryberg, J., & Tännsjö, T. (2014). The repugnant conclusion. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition). Stanford University.
plato.stanford.edu/archives/spr2014/entries/repugnant-conclusion

Ashford, E. & Mulgan, T. (2018). Contractualism. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Summer 2018 Edition). Stanford University.
plato.stanford.edu/archives/sum2018/entries/contractualism

Bartels, D. M. (2006). Proportion dominance: The generality and variability of favoring relative savings over absolute savings. *Organizational Behavior and Human Decision Processes, 100*(1), 76–95.
doi.org/10.1016/j.obhdp.2005.10.004

Baumann, T. (2017a). S-risk FAQ.
centerforreducingsuffering.org/research/faq

Baumann, T. (2017b). S-risks: An introduction.
centerforreducingsuffering.org/research/intro

Baumann, T. (2019/2022). Risk factors for s-risks.
centerforreducingsuffering.org/research/risk-factors-for-s-risks

Baumann, T. (2020a). Arguments for and against a focus on s-risks.
centerforreducingsuffering.org/research/arguments-for-and-against-a-focus-on-s-risks

Baumann, T. (2020b). Common ground for longtermists.
centerforreducingsuffering.org/research/common-ground-for-long-termists

Baumann, T. (2021). How can we reduce s-risks?
centerforreducingsuffering.org/research/how-can-we-reduce-s-risks

Baumann, T. (2022). *Avoiding the Worst: How to Prevent a Moral Catastrophe*.
centerforreducingsuffering.org/books/avoiding-the-worst-how-to-prevent-a-moral-catastrophe-by-tobias-baumann

Bostrom, N. & Ord, T. (2006). The Reversal Test: Eliminating Status Quo Bias in Applied Ethics. *Ethics, 116*(4), 656–679.
nickbostrom.com/ethics/statusquo.pdf

Brennan, S. M. K. (1988). *Ideal utilitarianism* (Doctoral dissertation). University of Iowa.
philpapers.org/rec/BREIU

Breyer, D. (2015). The cessation of suffering and Buddhist axiology. *Journal of Buddhist Ethics, 22*, 533–560.
blogs.dickinson.edu/buddhistethics/files/2015/12/Breyer-Axiology-final.pdf

Budolfson, M. & Spears, D. (2018). Why the repugnant conclusion is inescapable. *Princeton University Climate Futures Initiative working paper*. web.archive.org/web/20220614105405/https://scholar.princeton.edu/sites/default/files/cfi/files/budolfson_spears_2018_repugnant_cfi.pdf

Carlson, E. (1998). Mere addition and two trilemmas of population ethics. *Economics & Philosophy, 14*(2), 283–306. doi.org/10.1017/S0266267100003862

Carlson, E. (2007). Higher Values and Non-Archimedean Additivity. *Theoria, 73*(1), 3–27. doi.org/10.1111/j.1755-2567.2007.tb01185.x

Cavanna, A. E. et al. (2023). The Western origins of mindfulness therapy in ancient Rome. *Neurological Sciences*, *44*, 1861–1869. doi.org/10.1007/s10072-023-06651-w

Chappell, R. Y. & Meissner, D. (2023). Theories of Well-Being. utilitarianism.net/theories-of-wellbeing

Chappell, R. Y., Meissner, D., & MacAskill, W. (2021). Population Ethics: The Total View. utilitarianism.net/population-ethics#the-total-view

Chenoweth, E. & Stephan, M. J. (2011). *Why civil resistance works: The strategic logic of nonviolent conflict*. Columbia University Press. goodreads.com/en/book/show/10056014

Crisp, R. (2021). Well-being. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Winter 2021 Edition). Stanford University. plato.stanford.edu/archives/win2021/entries/well-being

Cudd, A. & Eftekhari, S. (2021). Contractarianism. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Winter 2021 Edition). Stanford University. plato.stanford.edu/archives/win2021/entries/contractarianism

Danaher, J. (2018). Popper's Critique of Utopianism and Defence of Negative Utilitarianism. philosophicaldisquisitions.blogspot.com/2018/01/poppers-critique-of-utopianism-and.html

Danaylov, N. (Host). (2013, October 15). Transhumanist Philosopher David Pearce: Give Up Eating Meat! [Video podcast episode]. youtube.com/watch?v=2wv8Hytp-c8

DeScioli, P. (2016). The side-taking hypothesis for moral judgment. *Current Opinion in Psychology*, *7*, 23–27. doi.org/10.1016/j.copsyc.2015.07.002

DiGiovanni, A. (2021a). A longtermist critique of "The expected value of extinction risk reduction is positive". forum.effectivealtruism.org/posts/RkPK8rWigSAybgGPe

DiGiovanni, A. (2021b). Tranquilism Respects Individual Desires.
anthonydigiovanni.substack.com/p/tranquilism-respects-individual-desires

DiGiovanni, A. (2023). Beginner's guide to reducing s-risks.
longtermrisk.org/beginners-guide-to-reducing-s-risks

Dodell-Feder, D. & Tamir, D. I. (2018). Fiction reading has a small positive impact on social cognition: A meta-analysis. *Journal of Experimental Psychology, 147*(11), 1713–1727.
doi.org/10.1037/xge0000395

Ewing, A. C. (1948). Utilitarianism. *Ethics*, *58*(2), 100–111.
doi.org/10.1086/290597

Fai. (2022). Why the expected numbers of farmed animals in the far future might be huge.
forum.effectivealtruism.org/posts/bfdc3MpsYEfDdvgtP

Fehige, C. (1998). A Pareto Principle for Possible People. In C. Fehige & U. Wessels (Eds.), *Preferences* (pp. 508–543). Walter de Gruyter.
fehige.info/pdf/A_Pareto_Principle_for_Possible_People.pdf

Fernando, J. W. et al. (2018). Functions of utopia: How utopian thinking motivates societal engagement. *Personality and Social Psychology Bulletin*, *44*(5), 779–792.
doi.org/10.1177/0146167217748604

Fletcher, G. (2016). *The Philosophy of Well-Being: An Introduction*. Routledge.
doi.org/10.4324/9781315745329

Fox, J. I. (2022). Does Schopenhauer accept any positive pleasures? *European Journal of Philosophy*.
doi.org/10.1111/ejop.12830

Frick, J. (2014). *'Making People Happy, Not Making Happy People': A Defense of the Asymmetry Intuition in Population Ethics* (Doctoral dissertation). Harvard University.
dash.harvard.edu/bitstream/handle/1/13064981/Frick_gsas.harvard.inactive_0084L_11842.pdf

Frick, J. (2020). Conditional Reasons and the Procreation Asymmetry. *Philosophical Perspectives, 34*(1), 53–87.
doi.org/10.1111/phpe.12139

Future of Life Institute. (2017). Superintelligence survey.
futureoflife.org/ai/superintelligence-survey

Gloor, L. (2016). The Case for Suffering-Focused Ethics.
longtermrisk.org/the-case-for-suffering-focused-ethics

Gloor, L. (2017). Tranquilism.
longtermrisk.org/tranquilism

Gloor, L. (2018). Cause prioritization for downside-focused value systems.
longtermrisk.org/cause-prioritization-downside-focused-value-systems

Gloor, L. (2022). Population Ethics Without Axiology: A Framework. forum.effectivealtruism.org/posts/dQvDxDMyueLyydHw4

Gómez-Emilsson, A. (2019). Logarithmic Scales of Pleasure and Pain. qri.org/blog/log-scales

Gómez-Emilsson, A. (2021). A Conversation with Roger Thisdell about Classical Enlightenment and Valence Structuralism. qri.org/blog/classical-enlightenment-and-valence-structuralism

Gómez-Emilsson, A. & Percy, C. (2023). The heavy-tailed valence hypothesis: The human capacity for vast variation in pleasure/pain and how to test it. *Frontiers in Psychology*, 14:1127221. doi.org/10.3389/fpsyg.2023.1127221

Goodman, C. (2016). Śāntideva. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Fall 2016 Edition). Stanford University. plato.stanford.edu/archives/fall2016/entries/shantideva

Gustafsson, J. E. (2022). Our Intuitive Grasp of the Repugnant Conclusion. In *The Oxford Handbook of Population Ethics* (p. 371). Oxford University Press. doi.org/10.1093/oxfordhb/9780190907686.013.25

Gyatso, T. (2002). *The Dalai Lama's Little Book of Wisdom*. Barnes & Noble. goodreads.com/book/show/418529

Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). Most people are not WEIRD. *Nature, 466*(7302), 29. doi.org/10.1038/466029a

Henrich, J., Heine, S. J., & Norenzayan, A. (2010b). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83. doi.org/10.1017/S0140525X0999152X

Hooker, B. (2023). Rule consequentialism. In E. N. Zalta & U. Nodelman (Eds.), *Stanford Encyclopedia of Philosophy* (Spring 2023 Edition). Stanford University. plato.stanford.edu/archives/spr2023/entries/consequentialism-rule

Hursthouse, R. & Pettigrove, G. (2023). Virtue ethics. In E. N. Zalta & U. Nodelman (Eds.), *Stanford Encyclopedia of Philosophy* (Fall 2023 Edition). Stanford University. plato.stanford.edu/archives/fall2023/entries/ethics-virtue

Häyry, M. (2024). Exit Duty Generator. *Cambridge Quarterly of Healthcare Ethics, 33*(2), 217–231. doi.org/10.1017/S096318012300004X

Jollimore, T. (2023). Impartiality. In E. N. Zalta & U. Nodelman (Eds.), *Stanford Encyclopedia of Philosophy* (Winter 2023 Edition). Stanford University. plato.stanford.edu/archives/win2023/entries/impartiality

Karlsen, D. S. (2013). Is God Our Benefactor? An Argument from Suffering. *Journal of Philosophy of Life, 3*, 145–167. philosophyoflife.org/jpl201309.pdf

Karnofsky, H. (2021). What counts as death?
cold-takes.com/what-counts-as-death

Knutsson, S. (2015). The 'Asymmetry' and extinction thought experiments.
simonknutsson.com/the-asymmetry-and-extinction-thought-experiments

Knutsson, S. (2016a). How Could an Empty World Be Better than a Populated One?
simonknutsson.com/how-could-an-empty-world-be-better-than-a-populated-one

Knutsson, S. (2016b). Measuring happiness and suffering.
simonknutsson.com/measuring-happiness-and-suffering

Knutsson, S. (2016c). Value lexicality.
simonknutsson.com/value-lexicality

Knutsson, S. (2016d). What is the difference between weak negative and non-negative ethical views?
simonknutsson.com/what-is-the-difference-between-weak-negative-and-non-negative-ethical-views

Knutsson, S. (2019). Epicurean Ideas about Pleasure, Pain, Good and Bad.
simonknutsson.com/epicurean-ideas-about-pleasure-pain-good-and-bad

Knutsson, S. (2021a). Many-valued logic and sequence arguments in value theory. *Synthese, 199*(3), 10793–10825.
doi.org/10.1007/s11229-021-03268-4

Knutsson, S. (2021b). The world destruction argument. *Inquiry, 64*(10), 1004–1023.
doi.org/10.1080/0020174X.2019.1658631

Knutsson. S. (2022a). Pessimism about the value of the future and the welfare of present and future beings based on their acts and traits.
simonknutsson.com/pessimism-value-future-welfare-acts-traits

Knutsson, S. (2022b). Undisturbedness as the hedonic ceiling.
simonknutsson.com/undisturbedness-as-the-hedonic-ceiling

Knutsson, S. (2023a). My moral view: Reducing suffering, 'how to be' as fundamental to morality, no positive value, cons of grand theory, and more.
centerforreducingsuffering.org/research/my-moral-view-reducing-suffering-by-simon-knutsson

Knutsson, S. (2024). Answers in population ethics were published long ago.
centerforreducingsuffering.org/research/answers-in-population-ethics-were-published-long-ago

Knutsson, S. & Thisdell, R. (2023). Roger Thisdell on undisturbedness, positive experiences, and the hedonic peak.
simonknutsson.com/roger-thisdell-on-undisturbedness-positive-experiences-and-the-hedonic-peak

Knutsson, S. & Vinding, M. (forthcoming). Introduction to suffering-focused ethics.

Kuhn, B. (2013). A critique of effective altruism.
forum.effectivealtruism.org/posts/mtj9NyXdnyhA8Qfw7

Kupfer, J. (2011). When the badness of vice outweighs the goodness of virtue: A moral asymmetry. *The Journal of Value Inquiry, 45*, 433–441.
doi.org/10.1007/s10790-011-9305-x

Laukkonen, R. E. et al. (2023). Cessations of consciousness in meditation: advancing a scientific understanding of nirodha samāpatti. *Progress in Brain Research, 280*, 61–87.
doi.org/10.1016/bs.pbr.2022.12.007

Lazari-Radek, K. & Singer, P. (2017). *Utilitarianism: A Very Short Introduction*. Oxford University Press.
doi.org/10.1093/actrade/9780198728795.001.0001

Leighton, J. (2017). Thriving in the Age of Factory Farming.
medium.com/@jonleighton1/thriving-in-the-age-of-factory-farming-fbcca7121d67

Leighton, J. (2023). *The Tango of Ethics: Intuition, Rationality and the Prevention of Suffering*. Imprint Academic.
goodreads.com/book/show/61889824

Leighton, J. (2024, March 20). Jonathan Leighton at EAGxAustralia 2023 - Unexpected Value: Prioritising the Urgency of Suffering [Video].
youtube.com/watch?v=4M5fgkWWujg

Lewis, G. (2016). Beware surprising and suspicious convergence.
forum.effectivealtruism.org/posts/omoZDu8ScNbot6kXS

MacAskill, W. (2022). *What We Owe the Future*. Basic Books.
goodreads.com/book/show/59802037

Markus, H. R. & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224–253.
doi.org/10.1037/0033-295X.98.2.224

Mayerfeld, J. (1999). *Suffering and Moral Responsibility*. Oxford University Press.
goodreads.com/book/show/2438988

Moorhouse, F., Plant, M., & Houlden, T. (2020). The philosophy of wellbeing.
happierlivesinstitute.org/report/the-philosophy-of-wellbeing

Mulgan, T. (2002). The reverse repugnant conclusion. *Utilitas, 14*(3), 360–364.
doi.org/10.1017/S0953820800003654

Mumper, M. L. & Gerrig, R. J. (2017). Leisure reading and social cognition: A meta-analysis. *Psychology of Aesthetics, Creativity, and the Arts, 11*(1), 109–120.
doi.org/10.1037/aca0000089

Nebel, J. M. (2022). Totalism without repugnance. In *Ethics and Existence: The Legacy of Derek Parfit*, 200–231.
doi.org/10.1093/oso/9780192894250.003.0009

Norcross, A. (2006). The scalar approach to utilitarianism. In H. R. West (Ed.), *The Blackwell Guide to Mill's Utilitarianism*, 217–232.
doi.org/10.1002/9780470776483.ch15

Ord, T. (2013). Why I'm Not a Negative Utilitarian.
amirrorclear.net/academic/ideas/negative-utilitarianism

Ord, T. (2015). Moral trade. *Ethics, 126*(1), 118–138.
fhi.ox.ac.uk/wp-content/uploads/moral-trade-1.pdf

Ord, T. (2020). *The Precipice: Existential Risk and the Future of Humanity*. Hachette Books.
goodreads.com/book/show/48570420

Orsi, F. (2012). David Ross, Ideal Utilitarianism, and the Intrinsic Value of Acts. *Journal for the History of Analytical Philosophy, 1*(2).
doi.org/10.4148/jhap.v1i2.1348

Parfit, D. (1984/1987). *Reasons and Persons*. Clarendon Press.
goodreads.com/book/show/327051

Pearce, D. (1995). *The Hedonistic Imperative*.
hedweb.com/hedab.htm

Pearce, D. (2005). The Pinprick Argument.
utilitarianism.com/pinprick-argument.html

Pearce, D. (2007). The Abolitionist Project.
abolitionist.com

Perry, L. (Host). (2022, October 6). Roger Thisdell on Centerlessness, Cessations, Defabrication, Valence, and Insight (No. 6) [Audio podcast episode].
youtube.com/watch?v=0EluGc-uf0Q

Popper, K. (1945). *The Open Society and Its Enemies*. Routledge. Wikiquote:
en.wikiquote.org/wiki/Karl_Popper#The_Open_Society_and_Its_Enemies_(1945)

Popper, K. (1947/1986). Utopia and Violence. *World Affairs, 149*, 3.
mudrac.ffzg.hr/~dpolsek/Popper_utopija%20i%20nasilje%20eng.pdf

Portmore, D. W. (1999). Does the total principle have any repugnant implications? *Ratio, 12*(1), 80–98.
doi.org/10.1111/1467-9329.00078

Prest, S. & Thisdell, R. (2022). Expansion & Contraction (the subtlest experiential distinction).
rogerthisdell.com/post/expansion-contraction-the-subtlest-experiential-distinction

Remmelt. (2021). Some blindspots in rationality and effective altruism.
forum.effectivealtruism.org/posts/LJwGdex4nn76iA8xy

Ridge, M. & McKeever, S. (2023). Moral particularism and moral generalism. In E. N. Zalta & U. Nodelman (Eds.), *Stanford Encyclopedia of Philosophy* (Summer 2023 Edition). Stanford University. plato.stanford.edu/archives/sum2023/entries/moral-particularism-generalism

Ryberg, J. (1996). Is the repugnant conclusion repugnant? *Philosophical Papers*, *25*(3), 161–177. doi.org/10.1080/05568649609506547

Ryder, R. (2001). *Painism: A Modern Morality*. Open Gate Press. goodreads.com/book/show/1283431

Schopenhauer, A. (1819/2012). *The World as Will and Idea (Vol. 3 of 3)*. Project Gutenberg. gutenberg.org/files/40868/40868-pdf.pdf

Schopenhauer, A. (1851/2004). *The Essays of Arthur Schopenhauer: Studies in Pessimism*. Project Gutenberg. gutenberg.org/files/10732/10732-h/10732-h.htm

Schroeder, M. (2018). Value theory. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Spring 2018 Edition). Stanford University. plato.stanford.edu/archives/spr2018/entries/value-theory

Seppälä, E. M. et al. (Eds.). (2017). *The Oxford Handbook of Compassion Science*. Oxford University Press. doi.org/10.1093/oxfordhb/9780190464684.001.0001

Shapiro, S. & Weisbaum, E. (2020). History of mindfulness and psychology. In *Oxford Research Encyclopedia of Psychology*. doi.org/10.1093/acrefore/9780190236557.013.678

Sherman, T. (2017). *Epicureanism: An Ancient Guide to Modern Wellbeing* (MPhil dissertation). University of Exeter. ore.exeter.ac.uk/repository/bitstream/handle/10871/32103/ShermanT.pdf

Singer, P. (1980). Right to Life? nybooks.com/articles/1980/08/14/right-to-life

Sinnott-Armstrong, W. (2023). Consequentialism. In E. N. Zalta & U. Nodelman (Eds.), *Stanford Encyclopedia of Philosophy* (Winter 2023 Edition). Stanford University. plato.stanford.edu/archives/win2023/entries/consequentialism

Smart, J. J. C. (1989). Negative utilitarianism. In *Freedom and Rationality* (pp. 35–46). Springer. doi.org/10.1007/978-94-009-2380-5_3

Spears, D. & Budolfson, M. (2021). Repugnant conclusions. *Social Choice and Welfare, 57*(3), 567–588. doi.org/10.1007/s00355-021-01321-2

Stanley, K. O. & Lehman, J. (2015). *Why Greatness Cannot Be Planned: The Myth of the Objective*. Springer. doi.org/10.1007/978-3-319-15524-1

St. Jules, M. (2019a). Conditional interests, asymmetries and EA priorities.
forum.effectivealtruism.org/posts/Sz7FrpdvSduQRSEnu

St. Jules, M. (2019b). Defending the procreation asymmetry with conditional interests.
forum.effectivealtruism.org/posts/2BWQ4NrCEP7a4vzaW

Thomas, T. (2018). Some possibilities in population axiology. *Mind, 127*(507), 807–832.
doi.org/10.1093/mind/fzx047

Thornley, E. (2022). A dilemma for lexical and Archimedean views in population axiology. *Economics & Philosophy, 38*(3), 395–415.
doi.org/10.1017/S0266267121000213

Todd, B. (2021). Expected value: how can we make a difference when we're uncertain what's true?
80000hours.org/articles/expected-value

Todd, B. (2023). Which jobs put you in the best long-term position?
80000hours.org/career-guide/career-capital

Tomasik, B. (2011). Risks of Astronomical Future Suffering.
longtermrisk.org/risks-of-astronomical-future-suffering

Tomasik, B. (2013a). Applied Welfare Biology and Why Wild-Animal Advocates Should Focus on Not Spreading Nature.
reducing-suffering.org/applied-welfare-biology-wild-animal-advocates-focus-spreading-nature

Tomasik, B. (2013b). Gains from Trade through Compromise.
longtermrisk.org/gains-from-trade-through-compromise

Tomasik, B. (2013c). How Would Catastrophic Risks Affect Prospects for Compromise?
longtermrisk.org/how-would-catastrophic-risks-affect-prospects-for-compromise

Tomasik, B. (2013d). Omelas and Space Colonization.
reducing-suffering.org/omelas-and-space-colonization

Tomasik, B. (2013e). Three Types of Negative Utilitarianism.
reducing-suffering.org/three-types-of-negative-utilitarianism

Tomasik, B. (2014a). Reasons to Be Nice to Other Value Systems.
longtermrisk.org/reasons-to-be-nice-to-other-value-systems

Tomasik, B. (2014b). Will Space Colonization Multiply Wild-Animal Suffering?
reducing-suffering.org/will-space-colonization-multiply-wild-animal-suffering

Tomasik, B. (2015a). Are Happiness and Suffering Symmetric?
reducing-suffering.org/happiness-suffering-symmetric

Tomasik, B. (2015b). Is Utilitarianism Too Demanding?
crucialconsiderations.org/ethics/demandingness

Tomasik, B. (2015c). A Small Mechanical Turk Survey on Ethics and Animal Welfare.
reducing-suffering.org/a-small-mechanical-turk-survey-on-ethics-and-animal-welfare

Tomasik, B. (2016). Strategic Considerations for Moral Antinatalists.
reducing-suffering.org/strategic-considerations-moral-antinatalists

Tännsjö, T. (2020). Why Derek Parfit had reasons to accept the Repugnant Conclusion. *Utilitas, 32*(4), 387–397.
doi.org/10.1017/S0953820820000102

van der Deijl, W. (2021). The sentience argument for experientialism about welfare. *Philosophical Studies*, *178*(1), 187–208.
doi.org/10.1007/s11098-020-01427-w

Vinding, M. (2015). *Anti-Natalism and the Future of Suffering: Why Negative Utilitarians Should Not Aim For Extinction*.
smashwords.com/books/view/543094

Vinding, M. (2017). Resources for Sustainable Activism.
magnusvinding.com/2017/12/30/resources-for-sustainable-activism

Vinding, M. (2020a). Clarifying lexical thresholds.
centerforreducingsuffering.org/research/clarifying-lexical-thresholds

Vinding, M. (2020b). Lexical views without abrupt breaks.
centerforreducingsuffering.org/research/lexical-views-without-abrupt-breaks

Vinding, M. (2020c). On purported positive goods "outweighing" suffering.
centerforreducingsuffering.org/research/on-purported-positive-goods-outweighing-suffering

Vinding, M. (2020d). *Suffering-Focused Ethics: Defense and Implications*. Ratio Ethica.
magnusvinding.files.wordpress.com/2020/05/suffering-focused-ethics.pdf

Vinding, M. (2020e). Why altruists should be cooperative.
centerforreducingsuffering.org/research/why-altruists-should-be-cooperative

Vinding, M. (2021). Comparing repugnant conclusions: Response to the "near-perfect paradise vs. small hell" objection.
centerforreducingsuffering.org/comparing-repugnant-conclusions

Vinding, M. (2022a). Comments on Mogensen's "The weight of suffering".
centerforreducingsuffering.org/comments-on-the-weight-of-suffering

Vinding, M. (2022b). Critique of MacAskill's "Is It Good to Make Happy People?".
forum.effectivealtruism.org/posts/vZ4kB8gpvkfHLfz8d

Vinding, M. (2022c). Lexicality between mild discomfort and unbearable suffering: A variety of possible views.
centerforreducingsuffering.org/lexicality-a-variety-of-possible-views

Vinding, M. (2022d). Lexical priority to extreme suffering — in practice.
centerforreducingsuffering.org/lexical-priority-in-practice

Vinding, M. (2022e). A phenomenological argument against a positive counter-part to suffering.
centerforreducingsuffering.org/phenomenological-argument

Vinding, M. (2022f). Point-by-point critique of Ord's "Why I'm Not a Nega-tive Utilitarian".
centerforreducingsuffering.org/point-by-point-critique-of-why-im-not-a-negative-utilitarian

Vinding, M. (2022g). *Reasoned Politics*. Ratio Ethica.
magnusvinding.files.wordpress.com/2022/03/reasoned-politics.pdf

Vinding, M. (2022h). Reply to Gustafsson's "Against Negative Utilitarianism".
centerforreducingsuffering.org/reply-to-gustafsson

Vinding, M. (2022i). Reply to the scope neglect objection against value lexical-ity.
centerforreducingsuffering.org/reply-to-the-scope-neglect-objection-against-value-lexicality

Vinding, M. (2022j). Some pitfalls of utilitarianism.
magnusvinding.com/2022/11/25/some-pitfalls-of-utilitarianism

Vinding, M. (2022k). A thought experiment that questions the moral im-portance of creating happy lives.
centerforreducingsuffering.org/a-thought-experiment-that-questions-the-moral-importance-of-creating-happy-lives

Vinding, M. (2022l). The dismal dismissal of suffering-focused views.
magnusvinding.com/2022/06/17/dismal-dismissal

Vinding, M. (2022m). Distrusting salience: Keeping unseen urgencies in mind.
magnusvinding.com/2022/12/08/distrusting-salience-keeping-unseen-urgencies-in-mind

Vinding, M. (2023). Minimalist versions of objective list theories of wellbeing.
magnusvinding.com/2023/04/10/minimalist-versions-of-objective-list-theories-of-wellbeing

Vinding, M. (forthcoming). *Compassionate Purpose: Personal Inspiration for a Better World*.
magnusvinding.com/books/#compassionate-purpose

Vinding, M. & Baumann, T. (2021). S-risk impact distribution is double-tailed.
centerforreducingsuffering.org/s-risk-impact-distribution-is-double-tailed

Whitehead, R. et al. (2018). Letting Go of Self: The Creation of the Nonattach-ment to Self Scale. *Frontiers in Psychology, 9*, 2544.
doi.org/10.3389/fpsyg.2018.02544

Wiblin, R. & Todd, B. (2018/2023). Ways people trying to do good acci-dentally make things worse, and how to avoid them.
80000hours.org/articles/accidental-harm

Wolf, C. (1996). Social Choice and Normative Population Theory: A Person Affecting Solution to Parfit's Mere Addition Paradox. *Philosophical Studies, 81*, 263–282.
doi.org/10.1007/bf00372786

Wolf, C. (1997). Person-Affecting Utilitarianism and Population Policy. In J. Heller & N. Fotion (Eds.), *Contingent Future Persons*. Kluwer Academic Publishers.
web.archive.org/web/20190410204154/https://jwcwolf.public.iastate.edu/Papers/jupe.htm

Wolf, C. (2004). O Repugnance, Where Is Thy Sting? In T. Tännsjö & J. Ryberg (Eds.), *The Repugnant Conclusion*. Kluwer Academic Publishers.
doi.org/10.1007/978-1-4020-2473-3_5

Woollard, F. & Howard-Snyder, F. (2022). Doing vs. allowing harm. In E. N. Zalta & U. Nodelman (Eds.), *Stanford Encyclopedia of Philosophy* (Winter 2022 Edition). Stanford University.
plato.stanford.edu/archives/win2022/entries/doing-allowing

Yu, S. et al. (2018). General need for autonomy and subjective well-being: A meta-analysis of studies in the US and East Asia. *Journal of Happiness Studies, 19*, 1863–1882.
doi.org/10.1007/s10902-017-9898-2

Yudkowsky, E. (2007). Torture vs. Dust Specks.
lesswrong.com/posts/3wYTFWY3LKQCnAptN

Zuber, S. et al. (2021) What should we agree on about the repugnant conclusion? *Utilitas, 33*(4), 379–383.
doi.org/10.1017/S095382082100011X