

On the Value of Happy People and the Moral Importance of Low-Welfare Species

Jakob Lohmar

St Anne's College
University of Oxford



A thesis submitted for the degree of
Doctor of Philosophy
25th April 2026

Acknowledgements

I would like to thank my supervisors Hilary Greaves, Andreas Mogensen, and Teruji Thomas for their generous support, great patience, and insightful feedback.

I am grateful to my college advisor Roger Crisp, who gave me additional support throughout my DPhil studies.

I have received helpful feedback for drafts of my chapters from more people than I can mention here. But I would like to give special thanks to Rhys Southan, Tomi Francis, Jonas Hertel, and Andrea Petrou – both for their feedback and their emotional support.

The Global Priorities Institute and Trajan House have provided me with the most productive and enjoyable workspace that I have ever had. I am grateful to everyone who contributed to making it such a special place.

I received generous financial support from the Effective Altruism Infrastructure Fund and the Heinrich Boell Foundation. I am very grateful for their support.

Table of Contents

Introduction.....	6
References.....	18
Chapter 1: Neutrality, Greediness, and Absurdity.....	20
1. Introduction.....	20
2. The Neutrality Intuition	24
3. The Greediness Phenomenon.....	28
4. Why Broome rejects Greedy Neutrality	31
4.1 The Dominance Problem.....	31
4.2 The Conceptual Problem.....	34
4.3 The Dependence Problem	36
5. The Welfare Problem.....	38
5.1 From the Greediness Phenomenon to the Welfare Problem	40
5.2 How the Welfare Problem relates to Impossibility Theorems	44
6. Conclusion	51
References.....	52
Chapter 2: In Defence of Pseudo-Dominance	56
1. Introduction.....	56
2. Weak and Strong Pseudo-Dominance.....	61
3. Objections from Incommensurateness	69
3.1 Paradigmatic Cases of Incommensurateness.....	70
3.2 Cyclicity	76
3.3 Crosswise Sweetening.....	78
4. Incommensurateness Requires Value Conflict	82
4.1 Weak and Strong Incommensurateness Requirements	82
4.2 An Argument for Weak Pseudo-Dominance.....	85
4.3 The Basic Case for the Weak Incommensurateness Requirement	87

4.4 Incommensurateness due to Ontological Differences	89
4.5 Incommensurateness due to Non-Existence of Determinates	92
4.6 Controversial Cases of Incommensurateness	95
5. The Application Problem	97
6. Solving the Application Problem	100
6.1 Ignoring Gerrymandered Respects	101
6.2 Ignoring Evaluatively Redundant Respects	103
6.3 Ignoring Outweighed Respects	112
7. How Pseudo-Dominance Refutes Neutrality	119
8. Conclusion	125
References	126
Chapter 3: Neutrality about Making Happy People and the Badness of Human Extinction	129
1. Introduction.....	129
2. Optimism about the Future of Humanity.....	133
3. No Badness despite Personal Harms	137
4. No Badness despite the Loss of Non-Welfarist Values.....	142
5. Conclusion	153
References.....	156
Chapter 4: The Moral Importance of Low-Welfare Species	159
1. Introduction.....	159
2. The Argument from Irrelevance of Tiny Benefits	163
3. The Problem of High-Welfare Species	166
4. The Argument from Fairness	170
5. The Argument from Empathy	172
6. Comparing Relative Magnitudes	173
7. Conclusion	177
References.....	177

Abstract

This thesis discusses two philosophical issues that are relevant to global priorities research: the value of additional happy people and the moral importance of species with small welfare ranges. Chapters 1 to 3 discuss the value of additional happy people, which is the focus of the thesis. The first chapter argues, based on John Broome's *greediness* argument, that a minimal welfarist requirement for the evaluation of variable-population cases can only be met by views on which additional happy people make the world better. Chapter 2 discusses the *pseudo-dominance* principle according to which something is better overall if it is better in some respect and not worse in any respect. It argues that a strict interpretation of this principle avoids problems stemming from incommensurateness but also makes the principle inapplicable. In response, two refined versions of pseudo-dominance are developed, and it is argued that these principles refute the neutrality view, according to which the existence of additional happy people is a neutral matter. Chapter 3 applies the neutrality view to the case of human extinction. It argues that even if we are extremely optimistic about the future development of humanity, near-term human extinction would not be bad if the neutrality view is true. Both the harms to present people and all non-welfarist values that could plausibly favour humanity's survival are cancelled out by the additional people in survival outcomes. There is thus no convergence of population-ethical views on the badness of human extinction. Chapter 4 finally turns to the moral importance of low-welfare species, i.e. species that have relatively small welfare ranges compared to humans. A natural extension of *relevance views* implies that all kinds of benefits to members of low-welfare species are irrelevant when significant human benefits are at stake. The chapter argues against this way of extending relevance views to inter-species comparisons and defends an alternative view, according to which even benefits to individuals with tiny welfare ranges can be relevant when much larger human benefits are at stake.

Introduction

We live in a world with scarce resources and many problems. Anyone who seeks to have a positive impact on the world faces the situation that even if they spent all their resources optimally, many problems would remain – there would still be a lot of suffering, injustice, and unfulfilled potential. Altruistic agents therefore need to prioritize some problems over others. The difficult question, which global priorities research investigates (see Greaves et al. 2020), is *which* problems we should prioritize. This thesis discusses two philosophical issues that are crucial for answering this question: the value of additional happy people and the moral importance of species with small welfare ranges. The thesis discusses these issues in four standalone papers that can be read independently of each other.

The focus of the thesis is on the value of additional happy people. The question is, roughly, whether the addition of people with good lives to a population makes the world better. To some, it is intuitively plausible that the answer to this question should be “yes” (e.g. Tännsjö 2002; Chappell 2017). But many others instead have the intuition that the existence of additional happy people is a *neutral* matter: their existence would make the world neither better nor worse (e.g. Narveson 1973, Frick 2017). Whether this “neutrality intuition”, as Broome (2004) calls it, is correct has drastic consequences for the question of prioritization. After all, most large-scale problems affect the expected size of the future population: global poverty causes many people to have more children than they otherwise would have; climate change will make parts of the earth uninhabitable while undermining food security in other parts of the earth; existential risks from nuclear war and unaligned artificial intelligence threaten to drastically reduce the size of future generations or to

eliminate humanity altogether.¹ That is not to say that it is obvious which total effects these and other large-scale problems will have on the future population size, even in expectation – there are many counteracting indirect effects and it is hard to say how they play out in the long run. But often there will be a strong effect on the expected future population size, and how this effect contributes to the overall value of the respective large-scale problem will crucially depend on whether the existence of additional happy people makes the world better.

In line with large parts of the literature on population ethics, I will talk about the value of additional happy *people*. But I believe that my arguments could be restated such that they refer more generally to happy *individuals* or *welfare subjects* without losing much (if any) plausibility. Given this generalisation, the relevance to the question of which problems an altruistic agent should prioritize becomes yet more ubiquitous. For example, factory farming requires more landmass than other forms of agriculture and thereby prevents the existence of many wild animals while, of course, causing the existence of many farmed animals. Many of these animals won't have lives that are worth living, but others will and whether we judge their existence to make the world better is crucial for how we evaluate certain forms of factory farming.

Besides for concrete problems that we could prioritize, the value of additional happy people is also relevant for more general views about how altruistic agents should use their resources. According to *longtermism*, we should act in ways that can be expected to be particularly good for the long-term future (e.g. Beckstead 2013, MacAskill 2022, Greaves & MacAskill 2025). Against what is sometimes assumed, longtermism does not imply that we should increase (or even maximize) the number of happy people in the future (see Wilkinson 2023). There are various ways of improving the long-term future that merely

¹ Of course, these are only two pertinent examples of the existential risks that humanity faces. See Ord (2020) for a detailed discussion of various existential risks.

aim at increasing the average quality of life and not the number of lives that will be lived in the long term. For example, one could mitigate the risks of a global totalitarian regime by strengthening democratic institutions (Caplan 2008), or one could work to improve societal values by preventing that bad values are locked-in through artificial intelligence (MacAskill 2022). However, if the existence of additional happy people makes the world better, many more promising long-term interventions emerge that strengthen the case for longtermism. In particular, the mitigation of risks of near-term human extinction would become extremely impactful since the prevention of human extinction adds in expectation astronomically many people to the world (Bostrom 2013; Greaves & MacAskill 2025, 25-27).²

Clearly, then, it matters a lot whether the existence of additional happy people makes the world better – or at least it matters a lot for anyone who has an interest in making the world better. At the end of the introduction, I will offer some reflections on whether the value of additional happy people also matters, or *should* matter, for people who don't have this interest. For now, I hope that the reader has at least a theoretical interest in the question and I will explain my approach to answering it in this thesis.

The basis of my arguments will be John Broome's well-known *greediness* argument against the neutrality intuition. According to this argument, the neutrality intuition commits us to the view that additions of happy people are not only neutral by themselves but also in conjunction with other good or bad things. As Broome (2005, 409) puts it, "we have found that our neutrality is *greedy*. Although neutral in itself, it is able to swallow up bad things and neutralize them". According to Broome, this makes the neutrality intuition implausible: "intuitively, neutrality should not behave like that; it should not be greedy" (ibid.).

² Another long-term intervention that becomes promising if additional happy people make the world better is space settlement. See Bostrom (2003) who argues that total utilitarianism implies we should maximise the probability that humanity ultimately settles space.

A nice feature of Broome's argument is that it does not presuppose that the neutrality intuition is part of any particular axiological view. Many critical discussions of the neutrality intuition consider different axiological views that entail the neutrality intuition and argue that all of them suffer from serious problems. For example, Greaves (2017, 8) argues that "it turns out to be remarkably difficult to formulate any remotely acceptable axiology that captures this idea of neutrality". Similarly, Ord (2020, 123) argues that the theories that entail the neutrality intuition "typically run into a variety of problems: conflicting with strong intuitions about thought experiments, conflicting with important moral principles, and violating widely held principles of rationality". If all theories that have been developed to justify the neutrality intuition suffer from serious problems, we should significantly lower our credence in the neutrality intuition. However, as far as these argument go, there remains the possibility of a yet undiscovered defensible theory that justifies – or is at least compatible with – the neutrality intuition. By contrast, Broome's greediness argument promises to refute the neutrality intuition directly, without making recourse to a more general axiology.

But many advocates of the neutrality intuition have not been swayed by Broome's argument. While they accept that neutrality is "greedy" in Broome's sense, they argue that neutrality's greediness is not such an implausible feature or that it is not a problem at all but rather a plausible implication of the intuition (e.g. Qizilbash 2007; Rabinowicz 2009; Frick 2017). I believe that these reactions to Broome's argument severely underestimate how problematic neutrality's greediness really is. The explanation for this underestimation seems to be that Broome does not focus on what is most implausible about neutrality's greediness as a problem *in its own right*, but only as an instance of more general problems. The first chapter of this thesis ("Neutrality, Greediness, and Absurdity") argues that once we isolate the most implausible feature of neutrality's greediness, we obtain a decisive argument against the neutrality intuition that is immune to the objections against Broome's

argument from the literature. What is most implausible about neutrality's greediness is that it leads to a view that fails to be minimally welfarist by implying that even immense welfare increases for many people are not good if they are accompanied by additions of happy people. I will call this problem the *welfare problem* and distinguish it from other problems with neutrality's greediness that Broome discusses. In particular, the welfare problem does not rely on the principle that "a good thing plus a neutral thing should add up to a good thing" (Broome 2005, 409), which has been rejected by several authors (e.g. Frick 2017; Hedden 2024). The welfare problem also does not rely on intuitions about the concept of neutrality or the intuition that we can generally ignore effects on the future population size when making value judgements, which have been called into doubt by other authors (e.g. Rabinowicz 2009). The only remaining defence of the neutrality intuition is to point out that *all* axiological views suffer from some counterintuitive implications, as impossibility theorems prove (e.g. Arrhenius 2000, 2003). In response to this defence, I will argue that the welfare problem goes beyond the intuitive costs that impossibility theorems force on every axiology. In particular, the welfare problem is worse than a mere violation of the principle known as "benign addition" (Huemer 2008) or "dominance addition" (Arrhenius 2000). Minimally welfarist views can reject this principle, but they cannot accept the welfare problem or the neutrality intuition.

One of the other problems with neutrality's greediness that Broome discusses is that it violates the principle that the combination of a good thing and a neutral thing is good overall. At some point, Broome (2005, 409) also uses a formulation of this principle in comparative terms, according to which something is better overall if it is better in one respect and not worse in the other respect. In chapter 1, I point out that principles of this kind have been described as *pseudo-dominance* principles and that they are typically rejected because they are incompatible with paradigmatic cases of incommensurateness and, relatedly, because they lead to betterness-cycles (e.g. Nebel 2019, 139; Hedden 2024).

However, I think that there is an interpretation of pseudo-dominance principles based on which a compelling answer to these objections can be given. I don't make this point in the first chapter because it requires a lot of theoretical argument. Instead, I have reserved chapter 2 ("In Defence of Pseudo-Dominance") for this theoretical deep dive.

The chapter argues that, if interpreted strictly, the simple pseudo-dominance principle, according to which something is overall better if it is better in some respect and not worse in any respect, is significantly weaker than the pseudo-dominance principles that have been refuted in the literature. These principles are – sometimes explicitly and sometimes implicitly – relativized to certain sets of evaluatively relevant respects, such as value dimensions or individual wellbeing levels. For example, *individual-wise* pseudo-dominance states that something is better overall if it is better with respect to some individual's wellbeing and not worse with respect to any individual's wellbeing. Even if the set of all individual wellbeing levels fully determines overall value, i.e. if *welfarism* is true, something can be worse in many respects even if it is not worse for any individual. For example, it could be worse with respect to total pleasure, or with respect to Amy's suffering. One might think that worseness in such respects cannot make a difference since, by assumption, overall value is fully determined by individual wellbeing levels. But in fact, worseness in those respects can make a crucial difference because it can lead to *incommensurateness* between the compared things. For example, even if one population is better than another for some individuals and not worse for any individual, the two populations could be incommensurate with respect to Amy's wellbeing *because* the first population is better with respect to Amy's pleasure while the second population is better with respect to Amy's suffering. By contrast, if the first population is not worse in any respect whatsoever – not just with respect to individual wellbeing levels but in any respect at all – this incommensurateness cannot arise because there are no conflicts between different welfare goods or any other values. Thus, the antecedent of the simple, non-

relativized pseudo-dominance principle rules out the above instance of incommensurateness.

I will argue that this is not a peculiarity of the above case. Within a suitably restricted domain of evaluands, all cases of incommensurateness are grounded in some kind of value conflict. This does not have to be a conflict between two different fundamental values, but each of the compared things needs to be better in at least some respect for incommensurateness to arise. I will show first that all cases in the literature that have been used to argue against pseudo-dominance principles include some value conflict in that weak sense. I will then generalize and argue for the view that, within a relevant domain of comparable things, each instance of incommensurateness is grounded in value conflicts. If this view is correct, the non-relativized pseudo-dominance principle is true.

Unfortunately, this is not the end of the story. The non-relativized principle is too weak to be applicable even in hypothetical cases. That is because, at least normally, betterness in some respect *necessitates* worseness in another respect. In response to this application problem, I will discuss several possible refinements of the simple principle that allow the pseudo-dominant thing to be worse in certain respects: *gerrymandered* respects, *evaluatively redundant* respects, and *outweighed* respects. I will argue that the latter two options lead to pseudo-dominance principles that are both applicable and defensible, in particular against objections from incommensurateness.

Finally, I will show how these two refined pseudo-dominance principles can be used to amend Broome's argument against the neutrality view. Importantly, besides the fact that both arguments presuppose the greediness phenomenon, this argument is independent of the argument given in the first chapter. The argument in the first chapter is based on welfarist intuitions, for which no pseudo-dominance principle is needed, and the argument in the second chapter is based on the application of certain pseudo-dominance principles, which does not presuppose any (non-trivial) welfarist intuitions.

Chapter 3 (“Neutrality about Making Happy People and the Badness of Human Extinction”) discusses the implications of the neutrality view for the value of human extinction. It argues that if the neutrality view is true, near-term human extinction would not be bad even under very optimistic assumptions about how humanity would develop in case of its survival. This conclusion might seem unsurprising: human extinction just means that there won’t be any future people who would have otherwise existed and the existence of these people does not have any positive value according to the neutrality view (cf. Broome 2022, 396). So how *could* human extinction be bad on the neutrality view?³ One simple answer is that an extinction event in the near future would most likely, at least, also severely harm all present (or near-future) people. Some proponents of the neutrality view have also attempted to give a more robust explanation for the badness of human extinction based on *non-welfarist* values (e.g. Frick 2017; Henning 2022, 143): humanity’s survival could be favoured by the value of humanity itself or by the value of scientific and cultural achievements that we could expect from future generations, for example. If we understand the neutrality view such that it only claims that additions of happy people don’t make the world better *in virtue of their happiness*, proponents of this view can consistently claim that human extinction would be bad with respect to certain non-welfarist values that future people would bring about. On the assumption that there are such non-welfarist values, and given that the last human generation would be severely harmed by human extinction, one would expect that even the neutrality view can account for the badness of human extinction. In fact, several authors who argue for the badness of human extinction have made this argument (e.g. Ord 2020, 124).

³ It is even natural to assume that human extinction would be *good* on the neutrality view since some future people would have bad lives that are not worth living, and the prevention of their existence seems clearly good (see Pummer 2020).

But the greediness phenomenon, once again, undermines what one would naturally expect from the neutrality view. The additional happy people in survival outcomes would cancel out whatever may be valuable about humanity's survival. Given the discussion in the first chapter, this is straightforward for the value of the welfare gains for present people. The focus of this chapter is therefore on non-welfarist values. It argues that the additional people in survival outcomes would cancel out even intuitively impressive non-welfarist values that might favour humanity's survival. Thus, if the neutrality view holds, human extinction would not be worse than even great survival outcomes that exceed the expectations of most optimists about the future of humanity.

Given the arguments in previous two chapters, one might wonder what the point of the argument in this chapter is: why should we care about the implications of the neutrality view if it is that implausible? In part, it provides yet another strong reason to reject the neutrality view. In fact, the argument in this chapter is structured as a *reductio*. This argument is not fully independent of the argument in the first chapter, but is also not redundant with it. It demonstrates the damage of neutrality's greediness in the intuitively especially severe case of human extinction, and – perhaps more importantly – it shows that the inclusion of non-welfarist values cannot help the neutrality view. I also think that we should take the result seriously that human extinction would not be bad if the neutrality view holds. Despite all of its serious problems, we should not be entirely certain that the neutrality view is false – especially given the impossibility theorems in population ethics – and we arguably should give some weight to any axiological view in which we have a positive credence when making important decisions about how to use large amounts of resources. There is also a reason to think that the neutrality view might be *especially* important to consider under moral uncertainty: it implies that human extinction is *incommensurate* with human survival. Just as under conditions of certainty, incommensurateness can also spread under uncertainty – it is “infectious”, as MacAskill

(2013) puts it. This means that even if we only have a small credence in the neutrality view, and all other views in which we have a positive credence imply that human extinction is worse than human survival, human extinction and human survival could be incommensurate overall relative to our credence distribution. This would be an unpalatable result, of course. I am hoping to investigate this issue in future research.

Chapter 4 (“The Moral Importance of Low-Welfare Species”) discusses whether benefits to members of species with small welfare ranges can, if numerous enough, outweigh significant benefits to humans. As I explain in this chapter, the question is important for ethical prioritization because it is plausible that if benefits to members of such low-welfare species can *in principle* outweigh significant human benefits, they *actually* outweigh them in some decision situations because of the fact that members of these species outnumber us by several orders of magnitude. To categorically avoid this practical consequence, it seems that we need to claim that their benefits are not even *relevant* when significant human benefits are at stake. Fortunately for us, this is exactly what a straightforward extension of partially aggregative views that employ the notion of relevance (e.g. Scanlon 1998, 239; Voorhoeve 2014) – in brief: “relevance views” – suggests: by definition, even the largest benefits to members of low-welfare species are very small compared to significant human benefits, and comparatively much smaller benefits are normally irrelevant to much larger benefits when all beneficiaries are humans. This chapter argues that – unfortunately for us – the relevance of benefits cannot be determined by comparing only their *absolute* magnitudes when the beneficiaries are members of species with significantly different welfare ranges. It is sufficient for a benefit’s relevance that its *relative* magnitude, which is determined in relation to the beneficiary’s welfare range (or the beneficiary’s *actual* welfare), is sufficiently large. Since the largest possible benefits for each beneficiary have a maximally large relative magnitude, even members of low-welfare species can be benefitted in ways that are always relevant.

Unlike in the previous chapters, the argument in chapter 4 is normative rather than axiological: it is about what we *ought to do* when we can benefit either humans or members of low-welfare species. The reason is that relevance views are only plausible as normative views – there are no axiologically irrelevant benefits. Since it is normative, the discussion in chapter 4 is clearly relevant to consequentialists and non-consequentialists alike. But it is less obvious whether the axiological discussion in the previous chapters is relevant to both consequentialists and non-consequentialists. The obvious thing to say in this context is that there is at least *some* moral reason to produce better outcomes rather than less good outcomes. I think this is indeed very plausible, and it is important to note that consequentialists and non-consequentialists can agree on this, even though some non-consequentialists in fact disagree (e.g. Sinclair 2018). But instead of stressing this familiar point, let me add a different conceptual perspective, which is both less common and, I think, more intuitive.

Clearly, what we *ideally should* do depends at least in part on axiological considerations: if one of two possible courses of action would have a better outcome, then we should, ideally, choose this course of action if all else is equal. I say “ideally” because otherwise there is the risk of understanding what we *should* do in terms of what we are *required* to do. It seems to me that there is a clear conceptual difference between the two: it is conceptually possible – and indeed quite plausible – that sometimes we *should* perform an action even though we are not *required* (or *obligated*) to perform it (e.g. because it is very demanding). But the risk of conflating the two concepts exists nevertheless because a lot of moral philosophy is concerned with deontic categories such as moral requirements and obligations. For that reason, claims about what we *ought to do* are often understood as claims about what we are *required* to do, and it seems plausible that if we *should* do something, then it is also the case that we *ought to do* it. Thus, it is easy to glide from claims about what we *should* do all the way to deontic claims about what we are required or

obligated to do. The “ideally” is there to stop this conceptual transition at the very beginning and to remind us that claims about what we should do are much weaker than claims about what we are required to do.

It seems plausible that what we ideally should do is determined by our *reasons* to perform certain actions rather than others. But we don’t have to make the detour via the concept of a reason to ascertain what we ideally should do. We can directly think about which features of our actions are relevant to what we ideally should do. At least sometimes, this way of thinking delivers the intuitively clearer results. Consider Bernard Williams’ (1995, 39) thought experiment of the callous husband who treats his wife badly. Given that “there is *nothing* in his motivational set that gives him a reason to be nicer to his wife as things are” (ibid.), it is questionable whether we can truly claim that the husband has a *reason* to be nicer to his wife. By contrast, it is obvious that he ideally *should* be nicer to his wife. Unlike Williams, I find plausible that the husband also has a *reason* to be nicer to his wife – despite his lack of concern for her. It is most plausible that it is both the case that he should be nicer to her *and* that he has a reason to be nicer to her. But it has also some plausibility that even though he ideally *should* be nicer to her, he unfortunately (due to his poor motivational profile) does not have any reason to be nicer to her. Either way, he clearly *should* be nicer to his wife.

I think that the same holds for performing acts with better outcomes (although not necessarily due to the possibility of reasons internalism). It is most plausible that it is both the case that we ideally should perform acts with better outcomes *and* that we have a reason to perform acts with better outcomes. But it also has some plausibility that we ideally should perform acts with better outcomes even though we don’t (always) have a reason to perform acts with better outcomes. Either way, if all else is equal, we clearly *should* ideally perform acts with better outcomes.

References

- Arrhenius, Gustaf. 2000. *Future generations: A challenge for moral theory*. Doctoral dissertation, Acta Universitatis Upsaliensis.
- . 2003. "The very repugnant conclusion." In *Logic, Law, Morality: Thirteen Essays in Practical Philosophy in Honour of Lennart Åqvist*, edited by Krister Segerberg and Rysiek Śliwiński. Uppsala: Uppsala University.
- Beckstead, Nicholas. 2013. *On the Overwhelming Importance of Shaping the Far Future*. PhD thesis, Rutgers University.
- Bostrom, Nick. 2003. "Astronomical Waste: The Opportunity Cost of Delayed Technological Development." *Utilitas* 15 (3): 308–314.
- . 2013. "Existential Risk Prevention as Global Priority." *Global Policy* 4 (1): 15–31.
- Broome, John. 2004. *Weighing Lives*. Oxford: Oxford University Press.
- . 2005. "Should We Value Population?" *Journal of Political Philosophy* 13 (4): 399–413.
- . 2022. "Climate Change and Population Ethics." In *The Oxford Handbook of Population Ethics*, edited by Gustaf Arrhenius et al. <https://doi.org/10.1093/oxfordhb/9780190907686.001.0001>.
- Caplan, Bryan. 2008. "The Totalitarian Threat." In *Global Catastrophic Risks*, edited by Nick Bostrom and Milan Ćirković. Oxford: Oxford University Press, 504–520.
- Chappell, Richard Y. 2017. "Rethinking the asymmetry." *Canadian Journal of Philosophy* 47 (2-3): 167–177.
- Frick, Johann. 2017. "On the survival of humanity." *Canadian Journal of Philosophy* 47 (2-3): 344–367. <http://dx.doi.org/10.1080/00455091.2017.1301764>.
- Greaves, Hilary. 2017. "Population axiology." *Philosophy Compass* 12 (11): e12442.
- , and William MacAskill. 2025. "The Case for Strong Longtermism." In *Essays on Longtermism*, edited by James Barrett, Hilary Greaves, and David Thorstad. Oxford: Oxford University Press.
- Greaves, Hilary, William MacAskill, Rossa O’Keeffe-O’Donovan, Philip Trammell, Benjamin Tereick, Andreas Mogensen, Christian Tarsney, Gustav Alexandrie, and Maxime Cugnon de Sévricourt. 2020. "A Research Agenda For The Global Priorities Institute." <https://globalprioritiesinstitute.org/wp-content/uploads/GPI-research-agenda-version-2.1.pdf>.

- Hedden, Brian. 2024. "Parity and pareto." *Philosophy and Phenomenological Research* 109 (2): 575–592.
- Henning, Tim. 2022. *Die Zukunft der Menschheit – soll es uns weiter geben?* Berlin: J.B. Metzler.
- MacAskill, William. 2013. "The infectiousness of nihilism." *Ethics* 123 (3): 508–520.
- . 2022. *What We Owe the Future*. London: Oneworld Publications.
- Narveson, Jan. 1973. "Moral problems of population." *The Monist* 57 (1): 62–86.
- Nebel, Jacob M. 2019. "Asymmetries in the Value of Existence." *Philosophical Perspectives* 33: 126–145.
- Pummer, Theron. 2020. "Review of *The Precipice: Existential Risk and the Future of Humanity*." *Notre Dame Philosophical Reviews*. <https://ndpr.nd.edu/reviews/the-precipice-existential-risk-and-the-future-of-humanity/>.
- Scanlon, Thomas M. 1998. *What We Owe to Each Other*. Cambridge, MA: Belknap Press of Harvard University Press.
- Sinclair, Thomas. 2018. "Are we conditionally obligated to be effective altruists?" *Philosophy & Public Affairs* 46 (1): 36–59.
- Tännsjö, Torbjörn. 2002. "Why We Ought to Accept the Repugnant Conclusion." *Utilitas* 14: 339–359.
- Voorhoeve, Alex. 2014. "How Should We Aggregate Competing Claims?" *Ethics* 125 (1): 64–87.
- Wilkinson, Hayden. 2023. "Can The Long Term Matter If Extinction Does Not?" GPI Working Paper. https://drive.google.com/file/d/18jqTY-glmgDuxo6TB-paDB7P4_7Bawni/view.
- Williams, Bernard. 1995. "Internal Reasons and the Obscurity of Blame." In *Making Sense of Humanity, and other philosophical papers*, 35–45. Cambridge, MA: Cambridge University Press.

Chapter 1: Neutrality, Greediness, and Absurdity

Abstract

According to the neutrality intuition, it is generally a neutral matter whether happy people come into existence. John Broome has shown that this intuition is committed to a *greedy* sort of neutrality: if the existence of happy people is neutral, their addition could cancel out other values. This greediness phenomenon has led Broome to abandon the neutrality intuition, but many others have not been swayed. This paper argues that Broome's reasons to reject the neutrality intuition as a consequence of the greediness phenomenon are not compelling because he does not consider what is most implausible about neutrality's greediness as a distinct problem in its own right: the greediness phenomenon shows that the neutrality intuition fails to be even minimally welfarist in variable population cases.

1. Introduction

Does the existence of additional people with good lives make the world better? The answer to this question crucially affects how we think about the current state of the world and various events that might alter it in the future: Is the world currently overpopulated or rather underpopulated? Is it a bad feature of climate change that it will diminish the inhabitable part of the earth? Is it desirable to spread the human population to other planets? Would human extinction in the near future be the most catastrophic event because it would prevent the existence of astronomically many future people?

Each of these questions has corresponding practical questions of major importance. Should we, for example, pursue policies that can be expected to reduce the population size

over time, or would this be a grave moral mistake? Should we make it our top priority to prevent human extinction, or should we rather focus on improving the lives of presently existing people in need? Given that we have at least a *pro tanto* reason to choose better over less good outcomes, questions like these cannot be fully answered without addressing the axiological question whether additions of happy people to the population make the world better.

Unfortunately, philosophers' intuitions dramatically come apart with regard to this question. On the one hand, many have the intuition that additions of happy people do make the world better. Huemer (2008), for example, employs the view that "it is better for there to be more lives with positive welfare" as a *premise* to which he seems to attribute at least some independent plausibility. More strongly, Tännsjö (2002), considers it "*obviously true*" that "the mere addition to the world of a person leading a life worth living makes the world better". On the other hand, there are also many (if not more) philosophers who have the intuition that additions of happy people normally don't make the world better.⁴ At least when the additional people are not *exceptionally* happy, their existence seems axiologically *neutral* to most of these philosophers.⁵ Perhaps the best-known example is Jan Narveson (1973, 80) who writes that "we are in favour of making people happy, but neutral about making happy people". Another well-known example is John Broome (2004) who dubs this "the intuition of neutrality".

However, Broome ultimately rejects the intuition of neutrality as a consequence of his *greediness* argument which shows that the intuition leads to the view that additions of

⁴ Interestingly, recent empirical research indicates that the intuition of neutrality is much less common among non-philosophers. In fact, lay people generally seem to be as much in favour of adding happy people to a population as they are against adding unhappy people (Caviola et al. 2022; see also Spears 2019 and Aaron's (2023) critique). Chappell (2017) argues that commonsense only supports that we have no moral obligation to procreate, but not that the existence of happy people is a neutral matter.

⁵ As discussed in the next section, some philosophers entertain the view that the addition of happy people could be *bad* in (at least) some circumstances (e.g. Benatar 2006, Frick 2022, Roberts 2022). However, this view is normally based on theoretical reasons and not considered intuitive in itself.

happy people would not only be neutral by themselves but also in conjunction with welfare increases or decrease for others. As Broome (2004, 170) puts it, the intuition is committed to “a sort of greedy neutrality, which is capable of swallowing up badness or goodness and neutralizing it”. This greediness *phenomenon* is relatively uncontroversial: if the neutrality intuition is correct, the neutrality at issue would need to be greedy in Broome’s sense.⁶ However, many advocates of the neutrality intuition disagree with Broome about whether the greediness phenomenon constitutes a decisive *problem* for the intuition. While Broome argues that it makes the intuition unacceptable, many others have argued that neutrality’s greediness is not such an implausible feature or that it is not a problem at all but rather a plausible implication of the intuition (e.g. Qizilbash 2007; Rabinowicz 2009; Frick 2017).

This paper argues that by demonstrating the greediness phenomenon, Broome has given us the resources for decisive arguments against the neutrality intuition. It is only that Broome doesn’t use these resources in the most compelling way. The most implausible aspect of neutrality’s greediness is that it commits us to the view that even immense welfare increases for many people are often not good if they are accompanied by additions of happy people. While Broome (2004, 204–5) notes this implication, he only discusses it in the context of the more general problem that all kinds of effects on people’s welfare could be neutralized by sufficient changes in population size. I will argue that it is crucial to consider the former implication not only as an instance of this more general *dependence problem*, but also as a problem in its own right, which I will call the *welfare problem*. This is important because the implication that welfare increases and decreases can be cancelled out by *some* changes in the number of happy people is neither particularly counterintuitive nor avoidable: *every* plausible axiology needs to accept that the value of welfare increases can

⁶ As we will see, a complication arises if the ‘better than’ relation is not transitive across different option-sets (see Frick 2022). I will argue that this restriction on transitivity is compatible with the relevant part of Broome’s argument.

be cancelled out by *reductions* in the number of happy people. Some advocates of the neutrality intuition have rejected Broome's argument for similar reasons and have consequently held on to the neutrality intuition (Rabinowicz 2009, 400; Frick 2017, 358). But none of their responses mitigate the welfare problem.

Another reason why Broome rejects the neutrality intuition is that the greediness phenomenon violates the principle that "a good thing plus a neutral thing should add up to a good thing" (2005, 409). Although this is not Broome's main argument, a lot of the discussion in the literature on neutrality's greediness has focused on it.⁷ The argument is commonly rejected because Broome's principle seems to be incompatible with incommensurateness (Rabinowicz 2009, 399; Frick 2017, 355; Hedden 2024, 581). A common lesson has been that neutrality's greediness is not an implausible feature, and that the neutrality intuition can be upheld despite the greediness phenomenon. But again, this is a mistake because what is most implausible about neutrality's greediness is the welfare problem, which does not rely on the above principle.

Some philosophers have recently discussed versions of the welfare problem, but without clearly distinguishing it from the other less serious problems with neutrality's greediness (Francis 2022, 68-69; Steele 2025, 163-64).⁸ One goal of this paper is to clearly distinguish the welfare problem from those other problems and to thereby establish it as a distinct problem. As noted above, this distinction is important because it reveals that the objections against the other greediness problems don't mitigate the welfare problem. On that basis, the paper continues to argue that there are some especially severe instances of the welfare problem that are unacceptable from a welfarist perspective; only a view that is

⁷ After presenting this argument, Broome writes: "This is implausible, but the full extent of its implausibility will not emerge till section 14.2, when we have some more theory worked out. Greedy neutrality will turn out to have some practical implications that are very implausible indeed" (2004, 170).

⁸ See also Adelstein (2024) who uses an adaptation of this argument to argue against a normative version of Neutrality according to which we have no moral reason to create happy people.

blind to the value of welfare could plausibly accept them. Finally, I will argue that these instances of the welfare problem are worse than violating some of the principles that are part of impossibility theorems. In particular, the welfare problem goes beyond a mere violation of the principle known as *Benign Addition*. Since the welfare problem is intuitively severe, it cannot be mitigated by extant defences of neutrality's greediness, and impossibility theorems don't force the problem on other views, I conclude that we should reject the neutrality intuition.

Here is the plan for the rest of the paper: After precisifying the relevant interpretation of the neutrality intuition (section 2) and reminding ourselves how Broome demonstrates the greediness *phenomenon* (section 3), I will discuss the reasons for which Broome takes the greediness phenomenon to refute the neutrality intuition (section 4). I will distinguish between three reasons, each corresponding to one greediness *problem*: the *dominance* problem, the *conceptual* problem, and the *dependence* problem. I will argue that, as they stand, none of these problems provides strong reason to reject the neutrality intuition. In contrast to these problems, I will introduce and discuss the *welfare* problem (section 5). I conclude that the welfare problem provides decisive reason to reject the neutrality intuition along with the related procreation asymmetry (section 6).

2. The Neutrality Intuition

Broome employs the greediness argument to argue against an axiological interpretation of the neutrality intuition. He suggests the following interpretation: "I interpret the neutrality intuition to be the view that the presence of an extra person in the world is neither good nor bad. More precisely: a world that contains an extra person is neither better nor worse than a world that does not contain her but is the same in other respects" (2005, 397). Broome later adds that this claim needs to be restricted to a certain *range* of wellbeing levels which

he calls the “neutral range” (ibid. 399): only if the wellbeing of the added people falls into this range is their addition neutral. The neutral range has a lower bound to exclude lives that are not worth living. Broome also allows that the neutral range could have an *upper* bound to exclude lives “at an extremely high standard of living” (ibid.) What would remain, in this case, is a “wide range of levels of lifetime well-being between a bad life and a very good life” (ibid.).

This view is significantly weaker than the view that most advocates of the neutrality intuition endorse. Most of them think that the existence of an additional happy person is not a good thing *no matter* how happy the person would be (e.g. Narveson 1973; Frick 2014; Bader 2022). This is the view that forms part of the (axiological) *Procreation Asymmetry* according to which it is bad to create miserable people but not good to create happy people.⁹ Nevertheless, I will follow Broome in discussing the weaker view that additions of happy people are neutral if their wellbeing falls into a certain range. I will only require that the neutral range is indeed a *wide* range that includes wellbeing levels that correspond to very good lives. To be more precise, I will assume that the neutral range extends at least to the wellbeing of the best-off presently existing people. This exact assumption is mostly for simplicity; we could also assume that there are some especially fortunate individuals whose wellbeing exceeds the neutral range. But at least the wellbeing of *most* well-off present people should fall into that range: the intuition is not that it is (in expectation) neutral to have children in a poor country with low living standards, but good to have children in a rich country with high living standards. Even having children in Scandinavia, where people seem to be the happiest, should normally be neutral rather than good if the neutrality intuition is true.¹⁰

⁹ Besides the above authors who endorse the Procreation Asymmetry, see also McMahan (1981, 2013) for a critical discussion.

¹⁰ It is worth mentioning that the welfare problem retains some of its force as long as the neutral range includes *some* positive wellbeing levels, no matter how narrow the range into which these levels fall. The only view

The neutrality intuition only concerns the non-instrumental value of additional people. It doesn't entail that additions of people in the neutral range are overall neutral if they have good or bad effects; the added people need to be *mere* additions. Broome therefore writes in the above quoted passage that the world with the added people must be "the same in other respects". This means in particular that all other people (and other welfare subjects) need to be equally well off in the two worlds. Since Broome (2004) restricts his analysis to the comparison of welfare distributions, which merely assign wellbeing levels to individuals, this is *all* that he requires to be equal (ibid. 146). I will follow suit but with a different terminology: I will compare states of affairs that are of the kind that certain *populations* exist, by which I mean sets of individuals with assigned wellbeing levels (i.e. the same as Broome's "distributions"). We can then formulate the axiological interpretation of the neutrality intuition as follows.¹¹

Neutrality: For all populations P and P^+ , where P^+ contains the people in P and additional people in the neutral range, P is neither better nor worse than P^+ if everyone who exists in both populations is as well off in P as in P^+ .

Neutrality leaves two possibilities for how P and P^+ compare: they could be *equally good* or they could be *incommensurate* in the sense that it is neither the case that they are equally good nor that one population is better than the other. Broome (2004) accordingly distinguishes between the equal goodness version and the incommensurateness version of

against which it loses most (if not all) of its force is the view that there is a neutral range that only includes lives that are personally neutral, i.e. neither good nor bad for the person living the life (see Rabinowicz 2022, Thornley 2022). This is good news for Rabinowicz who thinks that the greediness problem "is exacerbated given my reinterpretation of the neutral range" (ibid. 129). That is not true for the welfare problem which, as I will argue, is the main problem with greediness.

¹¹ As in the subsequent formulation, I will often just say that one population is better (worse, etc.) than another population. That is a convenient way of saying that the state of affairs that there is one population is better (worse, etc.) than the state of affairs that there is another population. I will also simplify the formulation by saying that people "are in the neutral range" if their wellbeing falls into that range.

Neutrality. I won't make this distinction since Broome's greediness argument applies against Neutrality in general; it does not assume a particular version of it.¹²

I will conduct my analysis in terms of Neutrality since it corresponds to Broome's axiological interpretation of the neutrality intuition. It is worth noting, however, that Neutrality is unnecessarily strong for my argument: the welfare problem (unlike the problems addressed by Broome) only require that additions of happy people are *not good*. Neutrality claims additionally that additions of people in the neutral range are *not bad* either. This is a respect in which Neutrality is stronger than the corresponding half of the Procreation Asymmetry: it's weaker by restricting its claim to a neutral range but it is stronger by claiming that additional people in this range are neutral rather than not good. One might think that this couldn't make a difference: it's clearly not *bad* to make happy people. However, some philosophers disagree. Infamously, Benatar (2006) holds that the existence of happy people is bad as long as their lives contain some suffering. Less radically, it has been argued that additions of happy people can be bad if they could have been added with a higher wellbeing instead, so that they are worse off than in this available alternative (Roberts 2011, 2022; Frick 2022). According to this view, it is never good but, under this condition, sometimes bad to add happy people. I will call this view *Conditional Neutrality* since it agrees with Neutrality under the extra condition that the additional people are not worse off than in an available alternative.

As I will show in the next section, the part of the greediness phenomenon that will be relevant for my arguments extends to Conditional Neutrality as well as all other views that imply the Procreation Asymmetry.

¹² Originally, Broome directed his greediness argument in particular against the incommensurateness version of Neutrality (2004, 169). This was sufficient for his purpose since he had already refuted the equal goodness version of Neutrality with a powerful argument (2004, 147). In later work (e.g. 2005, 2022), however, Broome employs the greediness argument against Neutrality in general.

3. The Greediness Phenomenon

According to Neutrality, mere additions of people in the neutral range result in populations that are neither better nor worse than the original population. Broome argues that if this claim was true, additions of people in the neutral range that make others worse off (and hence are not *mere* additions) would sometimes *also* result in populations that are not worse than the original population. Analogously, additions of people in the neutral range that make others *better* off would sometimes result in populations that are *not better* than the original population. These are the two parts of the *greediness phenomenon*, which we can state more precisely as follows.

Greediness Phenomenon: If Neutrality is true, then, for some populations P and some populations P+ containing the people in P and additional people in the neutral range, P is not better (worse) than P+ even though, among the people who exist in both populations, everyone is at least (at most) as well-off in P and some people are better-off (worse-off) in P.

Broome (2004, 170) demonstrates the first part of the Greediness Phenomenon with the following example, where A, B, and C are different populations and the numbers are wellbeing levels of the individuals in these populations. Numbers in one column refer to the wellbeing of the same individual (who accordingly exists in all populations in which a number is assigned to it) and “Ω” means that the individual does not exist in the respective population. All wellbeing levels are supposed to be in the neutral range.

$$A = (4, 6, \Omega)$$

$$B = (4, 6, 1)$$

$$C = (4, 4, 4)$$

Neutrality implies that A is not better (and also not worse) than B since B results from adding one person in the neutral range to A. Most plausibly, C is better than B: we get from B to C by increasing the wellbeing of the least well-off person by 3 units and decreasing the wellbeing of the most well-off person by only 2 units. By the transitivity of ‘better than’ (in short: *Transitivity*), it follows that A is not better than C if Neutrality is true.¹³ The only two differences between A and C are that C contains an additional person in the neutral range and that one person is better-off in A. Thus, if Neutrality is true, it is sometimes the case that a population P (such as A) is not better than another population P+ (such as C) that contains all people in P and additional people in the neutral range, even though some of the people who exist in both populations are better-off in P and the rest of them are equally well-off in both populations. This is the first part of the Greediness Phenomenon.

The second part can be demonstrated analogously. Here is one (self-chosen) example.¹⁴

$$D = (4, \Omega, \Omega)$$

$$E = (4, 6, 6)$$

$$F = (6, 4, 4)$$

By Neutrality, D is not worse (and also not better) than E since E results from adding two people in the neutral range to D. Most plausibly, F is worse than E: we can get from E to F

¹³ Assume that A is better than C. From this and the premise that C is better than B, it follows that A is better than B, which contradicts the implication of Neutrality that A is not better than B.

¹⁴ Broome mentions the second part of the Greediness Phenomenon but does not demonstrate it with the same level of detail with which he demonstrates the first part.

by swapping the wellbeing levels of two individuals and decreasing the wellbeing of another person from 6 to 4. By Transitivity, it follows that D is not worse than F if Neutrality is true. The only differences between D and F are that F contains two additional people in the neutral range and that one person is worse-off in D. Thus, if Neutrality is true, it is sometimes the case that a population P (such as D) is not worse than another population P+ (such as F) that contains all people in P and additional people in the neutral range, even though some of the people who exist in both populations are worse-off in P and the rest of them is equally well-off in both populations. This is the second part of the Greediness Phenomenon.

Importantly, the second part of the Greediness Phenomenon can be extended to all views that endorse the Procreation Asymmetry. That is because the relevant implication of Neutrality for demonstrating this part is that additions of happy people are *not good*. By definition, the Procreation Asymmetry agrees with this. In the above case, for example, Neutrality implies that E is not better than D. Since E differs from D only in that it contains additional happy people, the Procreation Asymmetry agrees with this verdict. As above, we can then argue that D is not worse than F if the Procreation Asymmetry is true.

Since Conditional Neutrality implies the Procreation Asymmetry, the second part of the Greediness Phenomenon can be extended to Conditional Neutrality. However, there is a complication. If Conditional Neutrality is true, the value of an option depends on the option-set from which it is chosen: whether mere additions of people are neutral depends on whether there is an available alternative (or “option”) in which these people are better off. Frick (2022) argues that, as a consequence of this option-set dependence, Transitivity only holds within option-sets but not across different option-sets. If this is right, we can only infer that D is not worse than F if the premises assume the same option-set as the

conclusion.¹⁵ To ensure that this is the case, we can stipulate that D, E, and F are the available alternatives in each case. We can then apply Transitivity to reach the conclusion that, on Conditional Neutrality, D is not worse than F if E is an available alternative.

4. Why Broome rejects Greedy Neutrality

Broome argues that Neutrality should be rejected as a consequence of the Greediness Phenomenon. He offers several reasons for this. I will distinguish between three different reasons, corresponding to (what I call) the *dominance* problem, the *conceptual* problem, and the *dependence* problem. The former two problems correspond to two different interpretations of one of Broome's arguments. I will comment on which of these Broome most plausibly had in mind, but I will focus on arguing that, at least without further argument, neither these nor the dependence problem provide strong reason to reject Neutrality.

4.1 The Dominance Problem

Right after establishing the Greediness Phenomenon, Broome usually introduces the principle that “a bad thing plus a neutral thing must add up to a bad thing” and, analogously, that “a good thing plus a neutral thing should add up to a good thing” (e.g. 2005, 409). The Greediness Phenomenon is supposed to reveal that Neutrality somehow conflicts with this principle. The question is how exactly this conflict is supposed to arise, which leads to two different interpretations.

¹⁵ In fact, I don't think that this is quite right. If Conditional Neutrality is true, Transitivity might not hold across different option-sets, as Frick argues. However, the compared populations don't need to be options of any agent; they could exist in possible worlds that are inaccessible to us (cf. Broome 2004, 147; Parfit 1984, 429–30). Hence, we only need Transitivity for comparisons in which there are *no* option-sets. Conditional Neutrality doesn't provide any reason to doubt Transitivity in such cases. Consequently, we only need to assume that the compared populations are not options to extend the Greediness Phenomenon to Conditional Neutrality.

According to one interpretation, Broome uses “neutral” and “neither good nor bad” interchangeably, so that his principle can be equally expressed by saying that a good (bad) thing and a thing that is *neither good nor bad* must add up to a good (bad) thing. At least in some of his works, this is indeed how Broome intends his principle to be understood. He unequivocally states in his (2005, 401) and (2022, 397) that “when I say something is ethically neutral, I shall always mean it is neither good nor bad”. Moreover, Broome offers a different (albeit implicit) formulation of his principle that does not employ the notion of neutrality at all and is directly stated in terms of better- and worseness: “So *C* is worse than *A* in one respect, and *neither better nor worse* in the other respect. Intuitively, therefore, *C* must on balance be worse than *A*” (2005, 409; 2022, 401; emphasis mine).

Thus understood, it is not hard to see how Broome’s principle conflicts with Neutrality given the Greediness Phenomenon. However, as Frick (2017, 355) points out, the principle is incompatible with (what many take to be) paradigm examples of incommensurateness. For example, it seems that each of two different careers (e.g. one as a musician and one as a painter) could be neither better nor worse than the other for some person both before and after a (slight) salary increase in one of the two careers. Accordingly, a change from one career to the other could be neither good nor bad for that person, the salary increase would be good for her, and still the combination of both changes could be neither good nor bad for her; the new career with the increased salary could be *incommensurate* with her original career.¹⁶ According to Broome’s principle this seems impossible: since the change in careers is neither good nor bad and the salary increase is good, the combination of both changes would need to be good overall.

There is another serious problem with Broome’s principle that comes out most clearly in his formulation of the principle in terms of better- and worseness in different

¹⁶ More specifically, the two careers could be *on a par* both before and after the salary increase (Chang 2002).

respects. Stated explicitly, his principle seems to be that, for any two things A and B that differ in two respects, if A is better than B in one respect, and neither better nor worse in the other respect, A is on balance better than B. Or, put more generally: for any two things A and B, if A is better than B in some respect, and neither better nor worse in any other respect, A is on balance better than B. Principles of this kind have been described as *pseudo-dominance* principles because they only require that A is *not worse* rather than *at least as good* in every respect, and they are typically rejected because they lead to cycles in certain cases involving incommensurateness: in these cases, A is better than B, B is better than C, and C is better than A (e.g. Nebel 2019, Hedden & Muñoz 2024).¹⁷ This is not only extremely implausible; Broome could also not accept this conclusion since he relies on the transitivity of ‘better than’.

Both arguments against Broome’s principle require that there are cases of incommensurateness – that sometimes neither of two things is at least as good as the other. That there are any such cases is not obvious. In fact, many have argued against the existence of incommensurateness, including Broome himself (1997; 2004, 174; see also Dorr, Nebel & Zuehl 2023; Tarsney, Lederman & Spears 2025). So it is not clear whether Broome’s principle actually fails to hold. However, a principle that only holds if there is no incommensurateness does not help in the present context. If there is no incommensurateness, the incommensurateness version of Neutrality is trivially false. To conclude this, we don’t need Broome’s principle. What remains is the equal goodness

¹⁷ Normally, this cyclicity is demonstrated by cases with three options that differ in three respects. One may wonder if a pseudo-dominance principle that is restricted to comparisons between options that differ in only *two* respects, which is all that Broome needs for his argument, can avoid cyclicity. As suggested in the main text, I think that such a principle seems plausible only as an instance of the more general principle. But we can also show that the restricted principle leads to cycles as well. Consider the following four options that differ in only two respects, where higher numbers are better and Ω is incommensurate with any number (the options can be interpreted as populations as elsewhere): A: (1, 3), B: (2, Ω), C: (3, 1), D: (Ω , 2). According to the restricted principle, B is better than A (since B is better in the first respect and incommensurate in the second respect), C is better than B (since C is better in the first respect and incommensurate in the second respect), D is better than C (since D is incommensurate in the first respect and better in the second respect), and A is better than D (since A is incommensurate in the first respect and better in the second respect).

version of Neutrality. We could use the principle to argue against this version. But we already know from Broome (2004, 147) that this version cannot plausibly be true for other reasons: it violates the “Principle of Personal Good” (i.e. Pareto). It seems, therefore, that Broome’s principle is either false (if there is incommensurateness) or does not contribute to showing that Neutrality is false (if there is no incommensurateness).¹⁸

4.2 The Conceptual Problem

There is an alternative interpretation of Broome’s principle that does not suffer from these problems. Perhaps, his principle really was supposed to be about the concept of *neutrality*, which might be more specific than neither good- nor badness. If we have an intuitive grasp of this concept, we could evaluate Broome’s principle independently of its version in terms of neither good- nor badness. The idea would be that it is part of our intuitive understanding of the concept of neutrality that the addition of a neutral thing to something good (bad) cannot result in something that isn’t good (bad). Whether this also holds for neither good- nor badness would then be a substantive question, not merely a terminological issue (cf. Gustafsson 2020, 109-10).¹⁹

So let us assume that Broome originally only endorsed the principle that “a bad thing combined with a neutral thing should be bad” (2004, 169), and analogously that a

¹⁸ There remains the possibility of turning things around by using Broome’s principle to argue against the existence of incommensurateness: Broome’s principle holds only if there is no incommensurateness. But Broome’s principle is really plausible. So we have a good reason to reject incommensurateness and with it the incommensurateness version of Neutrality. The problem with this argument is that it seems dialectically ineffective: those who are attracted to incommensurateness are often so *because* it allows us to say that neither of two things is better than the other both before and after certain (slight) improvements. But this is exactly what Broome’s principle rules out. That incommensurateness violates Broome’s principle will therefore hardly be considered a *problem* by those proponents of incommensurateness.

¹⁹ There is some textual evidence in Broome (2004) that this is indeed how he thought about the notion of neutrality when giving his argument from greediness. He writes that the Greediness Phenomenon teaches us that “incommensurateness simply does not capture accurately the intuitive idea of neutrality” (2004, 169) and that “incommensurateness is not neutrality as it intuitively should be” (170). These claims would be odd and close to self-contradictory if “neutrality” was just a placeholder for “neither good- nor badness”.

good thing combined with a neutral thing is good, without also endorsing the corresponding principle that replaces “neutral” by “neither good nor bad”. On this interpretation, Broome’s principle does not suffer from obvious counterexamples. But it is also unobvious whether the principle is true. One might think that we first need to get clarity on what “neutral” means before assessing Broome’s principle, and one might insist that “neutral” just means neither good nor bad, as Broome stipulates in some of his works, or that there is at least one relevant sense of “neutral” with that meaning (see Rabinowicz 2009, 399).

More importantly, if Broome’s principle only concerns the notion of neutrality, we lose the contradiction between it, the Greediness Phenomenon, and Neutrality. That is because Neutrality does not employ the notion of neutrality; it simply states that mere additions of people are *neither good nor bad*. We therefore *need* the conceptual link between neutrality and neither good- nor badness to contradict Neutrality with the help of the Greediness Phenomenon. Without this conceptual link, we can only conclude that mere additions of people are not *neutral*. This would make the naming of “Neutrality” inappropriate, but the principle itself could be maintained. What is more, Neutrality could also still be *motivated* in the usual ways. For one, many people simply have the intuition that adding happy people to a population is neither good nor bad. In fact, Broome later defines the neutrality intuition in this way and comments that he finds the intuition, thus understood, “very attractive” (2005, 401). Many others are attracted to Neutrality because it promises to avoid the conclusion that we have any *moral reason* to create additional people (e.g. Frick 2017, 348). To avoid this conclusion, we only need that additional people do not make the world better; we don’t *additionally* need that adding people is neutral. The same holds for other desiderata such as avoidance of the Repugnant Conclusion. I therefore conclude that Broome’s principle neither contradicts nor undermines Neutrality.

4.3 The Dependence Problem

Both the dominance problem and the conceptual problem are based on the principle that the combination of a neutral thing and a good (bad) thing must be good (bad) overall. But Broome does not only consider the Greediness Phenomenon a problem for Neutrality because it shows that Neutrality conflicts with that principle. He also considers it a problem because it shows that Neutrality conflicts with our common practice of ignoring effects on the future population size (in short: “population effects”) when making value judgements. In Broome’s (2022, 402) words: “The problem is not just that greedy neutrality is intuitively unsatisfactory. Much more serious is that, if neutrality is greedy, it cannot do the work we need from it. Remember it was the intuition of neutrality that apparently justified us in ignoring population effects when assessing some of the practical problems I described [e.g. climate change and road safety].” According to Broome, Neutrality does not allow us to evaluate these practical problems independently of their population effects since the latter might render them neutral despite their good or bad effects. I call this the *dependence* problem.

Given the Greediness Phenomenon, it is straightforward that Neutrality has this implication. The Greediness Phenomenon basically tells us that if Neutrality is true, the combination of changes in population size and welfare increases or decreases for others is sometimes not good or bad. But of course, welfare increases and decreases are good and bad, respectively, if they are not accompanied by population changes. Hence, whether harmful events that make people worse off are bad, and whether beneficial events that make people better off are good, depends on their population effects if Neutrality is true: they may turn out not good or bad overall if they alter the future population size. As Broome (2004, 203) argues, even events such as climate change that otherwise would be *extremely*

good or bad due to their effects on people's wellbeing could not be good or bad if they alter the population size sufficiently.

Is this implication a good reason to reject Neutrality? A first thing to note is that the dependence on population effects is not something that Neutrality is committed to *in particular*. After all, if Neutrality is false and additions of people *do* make a population better or worse unless they have one particular wellbeing level, effects on the future population size generally matter and cannot be ignored. It is Broome's insight that *even Neutrality* cannot avoid the verdict that population effects can alter the value of events. He therefore concludes that "our deeply-embedded practice of ignoring changes in population cannot be justified" (2009, 413-14).²⁰

Since this is a conclusion that we have to accept anyway, we cannot argue that Neutrality must be false because the dependence on population effects is absurd. It is not a *reductio* of Neutrality. Broome still considers the dependence on population effects a problem for Neutrality in particular since he considers avoidance of this dependence a crucial *motivation* for Neutrality. This creates an asymmetry with other axiological views, such as the total utilitarian view, that never had the ambition to avoid the dependence on population effects. However, as noted before, there are other reasons for which Neutrality has been endorsed, such as its coherence with the view that we have no moral reason to create happy people. For anyone motivated by these alternative reasons, the dependence problem undermines Neutrality at most partially.

I also think that independence of population effects should not be considered a desideratum in the first place. Even events that otherwise would be extremely good or bad could intuitively cease to be good or bad if they have certain population effects. Broome

²⁰ As Rabinowicz (2009, 400) argues, and as Broome recognizes, even Broome's preferred vagueness view, on which it is *indeterminate* whether mere additions of people make populations better or worse, cannot avoid this conclusion.

(e.g. 2022, 402) argues for the opposite based on the example of climate change which he considers clearly bad simply because of all the harm that it will do to millions of people. He considers it implausible that changes in population size could cancel out this harm. But assume that for every person that climate change harms, it enables the existence of one million future people with lives that are well worth living. If you want, assume additionally that without climate change each future generation would only contain a small number of people. Under that assumption, would climate change still be bad? I do not find that intuitively compelling. Indeed, I find intuitive that climate change would *not* be bad overall if it had these population effects.²¹

5. The Welfare Problem

I have argued that, as they stand, none of the greediness problems that Broome discusses provide strong reason to reject Neutrality. But the discussion of Broome's arguments suggests a more serious problem. In response to the dependence problem, I have argued that it is not always unintuitive that harmful events could cease to be bad if they also enable the existence of many people with good lives. However, the reverse claim that harmful events can cease to be bad if they also *prevent* the existence of happy people seems implausible indeed. Similarly, it seems implausible that beneficial events can cease to be *good* if they also *add* happy people to the population. But the Greediness Phenomenon tells us that this is sometimes the case according to Neutrality: the value of welfare increases is sometimes cancelled out by additions of happy people. That seems implausible. What is more, unlike dependence on population effects in general, this particular implication can

²¹ Perhaps Broome disagrees with this because he rejects the notion of a life worth living, or of a life that is good for the person living it (see 2004, 68; 2007, 119-20): it seems indeed implausible that climate change could not be bad merely because it increases the number of future people if we could not meaningfully say that these people would have good lives. Perhaps, this also explains why Broome doesn't consider the welfare problem, which I will turn to next, as a problem in its own right.

be avoided by views that don't imply Neutrality. For example, it is always good to make happy people while making others better off according to a total utilitarian view.

But perhaps we should not be very confident that this verdict is correct. After all, the discussion of the dominance problem and of the conceptual problem shows that we should not rely on the principle that the combination of a good thing and a neutral thing is good overall. Hence, we cannot argue that the combination of welfare increases and additions of happy people *must* be good because the welfare increases are good while the addition of happy people is not bad. Instead, we have to find a case in which this kind of verdict is intuitively compelling in itself. Arguably, it is not *always* compelling, such as when the welfare increases are minimal while the lives of the added people are only barely worth living. But I will argue that there are other cases in which, intuitively, only a view that is blind to the value of welfare could plausibly deny that the combination of welfare increases and additions of happy people is good. That Neutrality cannot agree with the intuitive verdict in such cases constitutes the *welfare problem*.

Since the welfare problem does not rely on the principle that the combination of a good thing and a neutral thing is good overall, it is easily distinguishable from the dominance problem and the conceptual problem. The difference between the welfare problem and the dependence problem is more subtle, so it's worth being explicit about it. Both problems point out practical implications of Neutrality's greediness that are supposed to be implausible in themselves. However, while the dependence problem rejects the implication that changes in population size can make an important difference to overall value, the welfare problem only insists that they cannot affect overall value in a specific way: at least when the welfarist stakes are especially high (in the ways specified in the next subsection), the addition of happy people cannot cancel out the positive value of welfare increases. The welfare problem doesn't rely on the intuition that population size is an unimportant factor that can be ignored. We may think that population effects can make a

crucial difference (which, as argued, we need to accept anyway) but that the difference that they make cannot be of the anti-welfarist kind that the welfare problem rejects. This pair of intuitions is natural for anyone who is not attracted to the neutrality intuition in the first place. But even advocates of Neutrality may share these intuitions. Frick seems to be one example. He writes that the intuition of neutrality is captured by a “holistic” principle according to which additions of happy people are neutral even in combination with some welfare *decreases* for others (Frick 2017, 358). He doesn’t say the same about welfare increases. Presumably, this is because, even to Frick, the latter seems unintuitive.²² Of course, one does not have to go as far as Frick and consider it positively intuitive that population increases can cancel out welfare decreases. One could find this intuitively unclear while finding counterintuitive that the same holds for welfare increases. Either way, it is important to consider the welfare problem as a problem in its own right, not merely as an instance of the dependence problem.

5.1 From the Greediness Phenomenon to the Welfare Problem

Consider again the case that I used for demonstrating the second part of the Greediness Problem. The result was that, on Neutrality, population F is not better than population D.²³

$$D = (4, \Omega, \Omega)$$

$$F = (6, 4, 4)$$

²² In another paper, Frick defends that even the combination of additional happy people and welfare increases is sometimes not good. But that is not because he considers this intuitive but because of “how extremely unappealing the alternatives are” (Frick 2022, 254). Considering his welfarist concerns (see the next subsection), Rabinowicz should consider the welfare problem more serious than the dependence problem as well.

²³ See Broome (2004, 204) for a similar case.

That seems implausible: F differs from D in that, first, it contains two additional people with lives worth living and in that, second, one person is better off in F than in D. But we can make Neutrality's implication intuitively much worse by interpreting the case as follows. First, assume that the numbers in each column represent the wellbeing of not one person but very many – say one billion – people. Second, assume that a wellbeing of 4 corresponds to a life that is *well* worth living (that is: more than barely worth living) and that a life at wellbeing 6 is significantly better yet. As long as the latter wellbeing level still falls into the neutral range, this assumption is innocuous; Neutrality still implies that F is not better than D. Of course, if the neutral range has no upper bound, wellbeing 4 and 6 could be arbitrarily high and their difference arbitrarily great. If the neutral range does have an upper bound, we can assume that wellbeing 6 represents the highest wellbeing in the neutral range. Given my assumption that this wellbeing is at least as high as the wellbeing of the best-off present people, a significantly lower wellbeing, such as 4, could still correspond to lives that are well worth living. For example, the lives of many well-off present people would plausibly still be well worth living if they included a painful tropical disease, or one year of depression, or one hour of torture. Wellbeing 4 would then represent this reduced but still overall good wellbeing. Finally, let us assume that the people who exist in both populations are worse off in D *because* of additional suffering that they are spared from in F. As argued above, this suffering could amount *at least* to one year of depression (etc.), but it could also be much more if the neutral range has a higher or no upper bound. At the same time, we can assume that the lives of the additional people in F don't contain this additional suffering. In fact, we could assume that their lives don't contain *any* suffering: the reason why they are only as well off as the people in D, whose lives contain the additional suffering, is simply that they have less of the good things in life.

With all of these assumptions in place, could F not be better than D, as Neutrality implies? The one billion people who exist in both populations endure some significant

suffering in D that they are spared from in F. These people would still have overall good lives in D, but they would all much prefer to live in F with less suffering. The only other difference between D and F is that F contains two billion additional people with good lives. These lives are not as good as the lives of the other people in F, but they are still well worth living and no worse than the lives of the people in D; their lives also don't contain any suffering. Overall, these differences seem to make F much better than D.²⁴

Perhaps more impressively, we can think of an event that threatens to change the world's population from F to D: it would inflict suffering on all present people while preventing the existence of many future people who would have good lives. Broome's (2004, 204) example of climate change is fitting here: unmitigated climate change would severely harm present people and prevent the existence of many future people by rendering parts of the earth uninhabitable. I argued before that sufficiently many additions of happy people might be able to cancel out the harms that climate change inflicts on present people. But the harms of climate change certainly cannot be cancelled out by *reductions* of the number of future happy people. Given that unmitigated climate change would have this population effect, there is no doubt that it would make the world overall much worse. Again, Neutrality cannot agree with this verdict.

I have stipulated above that the people who live in both populations are better off in F because they suffer more in D, and that the lives of the additional people in F don't contain suffering. These stipulations add intuitive force to the case and, of course, make the verdict that F is better than D especially hard to deny on suffering-focused views.²⁵ From a

²⁴ As noted in the context of Conditional Neutrality, we don't need to assume that D and F are available alternatives to any agent. Still, we can ask: does our intuition change if we assume that D and F are available alternatives besides a third available outcome E in which the additional people are better off than in F? I don't think it does: even if the additional people in F could have been made much better off, the addition of their still well worth-living lives in combination with the immense suffering reduction for one billion other people is overall good.

²⁵ For example, Benatar (2006) only considers the addition of people to be bad if their lives contain suffering – and of course he is in favour of reducing the suffering of other people. More generally, anyone who prefers populations with less total and average suffering should prefer F over D.

welfarist perspective, according to which what ultimately matters is the wellbeing level of each individual, these stipulations are unnecessary – they only help to illustrate the welfarist stakes. What matters from such a welfarist perspective is that the people who exist in D are much better off in F (whether or not this wellbeing difference is grounded in different amounts of suffering) and that the additional people in F have lives that are both well worth living and no worse than the lives of the people in D (whatever goods and bads are responsible for this wellbeing). It is worth isolating these purely welfarist differences between F and D. Jointly, they seem sufficient to make F clearly better than D. A view that denies this verdict violates what seems to be a minimal welfarist principle for the evaluation of variable population cases.

Minimal Welfarist Population Principle

A bigger population P^+ is better than a smaller population P if everyone who exists in P also exists in P^+ but with a significantly higher wellbeing and everyone who exists in P^+ but not in P is well off in absolute terms and also at least as well off as everyone in P .²⁶

In a recent paper, Rabinowicz notes that according to Neutrality there is a “striking *disparity* between what is good for a person and what is impersonally good: a life might be good for the person who lives it without being impersonally contributively good—without making the world better” (2022, 118). As he continues to argue, this disparity might be worrying from a welfarist perspective, but it barely provides an argument against Neutrality since Neutrality explicitly *endorses* this disparity. What my argument shows is that, surprisingly,

²⁶ That the people who exist in P^+ but not in P are “well off” is supposed to capture that their lives are *well* worth living. We could also add another condition to the principle: it holds for *sufficiently large* populations P , so that all these people benefit from the welfare increases in P^+ .

Neutrality is committed to a much more extreme disparity between personal and impersonal value. The problem is not that Neutrality rejects some straightforward extension of welfarism to variable population cases. The problem is that Neutrality fails to meet a minimal welfarist requirement for the evaluation of such cases by violating the Minimal Welfarist Population Principle. This should make Neutrality hardly acceptable to anyone with a broadly welfarist disposition; only a view that does not recognize the value of welfare could plausibly deny this principle.

5.2 How the Welfare Problem relates to Impossibility Theorems

The welfare problem is intuitively severe and can be avoided by competing views. It thereby threatens to be a *reductio* of Neutrality. To succeed as a *reductio*, it needs to be shown that the welfare problem is worse than the implausible implications of other views. Given the impossibility theorems in population ethics (e.g. Arrhenius 2000a), which show that *every* population axiology violates some intuitively compelling principle, this is a high bar. But I will argue that the welfare problem surpasses this bar. It is worse than a mere violation of one of the principles that impossibility theorems prove to be incompatible.

There is a principle which figures in some impossibility theorems that is closely related to the welfare problem.²⁷ According to *Benign Addition* (Huemer 2008) – also called *Dominance Addition* (Arrhenius 2000a) – additions of happy people that are accompanied by welfare increases for everyone else – call them “benign additions” – make a population better. The welfare problem shows that Neutrality violates a logically weaker and

²⁷ In particular, these are Arrhenius’ second (2000a) and fifth (2003) impossibility theorems. See also the works from other authors referenced in footnote 24 for some arguments and paradoxes that are closely related to Arrhenius’ second impossibility theorem.

intuitively more compelling version of Benign Addition. The Minimal Welfarist Population Principle states that benign additions are good *if* they have the following desirable features.

- (i) The lives of the added people are *well* worth living.
- (ii) The added people are at least as well off as the original people were before their addition.
- (iii) The welfare of the original people increases *significantly*.

The welfare problem entails that even benign additions with these features are not always good according to Neutrality. As we have seen, this even holds if benign additions have additionally the following features.

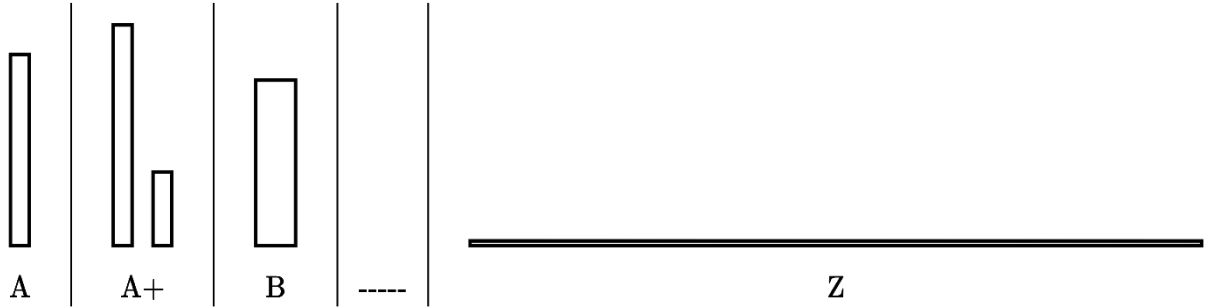
- (iv) The welfare of the original people increases as a consequence of reduced suffering.
- (v) The lives of the additional people don't contain any suffering.

For a direct comparison with Benign Addition, let's say that *Superb Addition* states that benign additions are good if they have these five features.²⁸

Importantly, these additional conditions make Superb Addition too weak to fill the role that Benign Addition plays in impossibility theorems. Generally speaking, the role of Benign Addition is to show jointly with some other principles that every population of people with great lives can be transformed into a population of people with lives that are barely worth living (or, in part, even worth *not* living) via a sequence of improvements. By Transitivity, this allows inferring the Repugnant Conclusion (or some version of it) according to which there is for every population of people with great lives a better population that only contains people with lives that are barely worth living. For illustration,

²⁸ Compare with Wilkinson's (2023, 23) "Excellent Benign Addition" according to which benign additions are good if the added people have a sufficiently high wellbeing. Although not strictly weaker (since even people with *well* worth-living lives might not be *sufficiently* well off), Superb Addition is yet more plausible than this principle. We could also apply a strictly weaker version which adopts the condition of a sufficiently high wellbeing to argue against the unrestricted version of Neutrality. A similar remark applies to Adelstein's (2024, 342) "Extreme Generative Improvement".

consider the following block diagram which represents one part of the sequence of the by now canonical version of Parfit's (1984) Mere Addition Paradox, which replaces the principle of Mere Addition by Benign Addition.²⁹ (The width of the blocks represents the population size, and the height of the blocks represents the population's wellbeing level.)

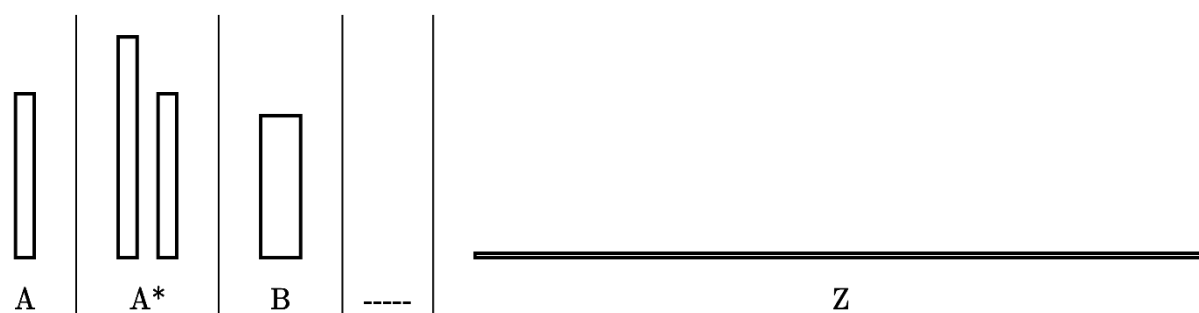


The role of Benign Addition with regard to this part of the sequence is to justify the premise that A+ is better than A, where A is a population of people with great lives and A+ is a population of the same people with yet higher wellbeing (represented by the left box) plus additional people with lower but still positive wellbeing (represented by the right box). Importantly, the wellbeing of the additional people in A+ is lower to such an extent that the average wellbeing in A+ is lower than in A. Nevertheless, since A+ results from A by adding happy people while making the original people better off, Benign Addition implies that A+ is better than A. This forms the first part of a sequence of improvements that leads to population Z in which everyone has a life that is barely worth living. Next, we add the premise that B, which results from redistributing the wellbeing of the people in A+ equally while adding some extra wellbeing, is better than A+. This follows, for example, from *Non-Anti-Egalitarianism* according to which a perfectly equal population is better than any population with the same people and lower total (and hence average) wellbeing (Ng 1989).

²⁹ Arrhenius (2022) calls this the “Sequential Dominance Addition Paradox”. There are several related ways in which it has been shown that Benign Addition leads to the Repugnant Conclusion when conjoined with other plausible principles. See, for example, Huemer (2008), Tännsjö (2002), and Parfit (1984, 438). The role of Benign Addition is, however, basically the same in all of these versions.

By Transitivity, it follows that B – a larger population with lower average wellbeing – is better than A. With the same principles, it can analogously be argued that a yet larger population C with yet lower average wellbeing is better than population B, and so on, until we reach an enormous population Z containing only people with lives barely worth living. By Transitivity, it follows that Z is better than A – the Repugnant Conclusion. If we accept Benign Addition, we therefore need to reject Non-Anti-Egalitarianism or Transitivity or accept the Repugnant Conclusion.³⁰

The most striking reason why Superb Addition cannot play the role of Benign Addition is that the additional people in A+ are worse off than the people in A. Since clause (ii) of Superb Addition requires that the additional people are at least as well off as those people, Superb Addition does not imply that A+ is better than A. The same holds for all other parts of the sequence in which Benign Addition is applied. For illustration, here is how population A would need to be changed to be better according to Superb Addition.



Superb Addition implies that A* is better than A. But the transition to A* does not move us any closer to population Z. In particular, Non-Anti-Egalitarianism does not imply that B is better than A* since B has lower average wellbeing than A*. On the contrary, B is clearly *worse* than A* since everyone is worse off in B than in A*.

³⁰ Another possibility is to reject the implicit assumption that we can get from the great lives in A to the drab lives in Z by a finite number of small decreases of their quality (see Thomas 2018).

Some of the other conditions of Superb Addition provide further reasons why Superb Addition cannot play the role of Benign Addition. Condition (i) requires that the additional people have lives that are *well* worth living. This may be the case for the additional people in A+, but at *some* point in the sequence the added people must have lives that are only barely worth living. Otherwise, we could not argue (via Non-Anti-Egalitarianism) that population Z in which *everyone* has a life that is barely worth living is better than the preceding population. Finally, consider condition (iii) in combination with condition (iv). They jointly require that the people who exist in both A and A+ suffer significantly less in A+ than in A. That means that the lives of the people in A, which are supposed to be great, need to contain significant suffering. Arguably this is possible but makes the conclusion that Z is better than A less repugnant. Either way, it could not be shown for *any* population A of great lives that Z is better than A, as the Repugnant Conclusion states.³¹

Since Superb Addition is weaker than Benign Addition and (consequently) cannot play its role in impossibility theorems, the welfare problem confers additional costs on Neutrality that go beyond the costs that impossibility theorems force on every axiology: without a commitment to Neutrality, we could merely violate Benign Addition or some of the other principles that are incompatible with it while upholding Superb Addition. For anyone who considers some of the other principles less compelling than (even) Benign Addition, part of the additional cost for Neutrality is simply that it violates Benign Addition rather than one of the less compelling principles.³² But even if one considers Benign

³¹ Condition (iv) also creates trouble in combination with condition (v). The latter requires that the additional people in A+ don't suffer while (iv) requires that their suffering is later reduced in moving from B to B+. This is only possible if we add suffering to their lives in moving from A+ to B, which makes the claim that B is better than A+ less appealing.

³² It has been argued, for example, that accepting the Repugnant Conclusion is less implausible (e.g. Tännsjö 2002; Huemer 2008).

Addition the weakest link in impossibility theorems, there remains the cost that Neutrality gratuitously violates the weaker principle of Superb Addition. I have pointed out before that Superb Addition is intuitively more compelling due to the additional conditions that it posits. These conditions also ground principled reasons to uphold Superb Addition even if one rejects Benign Addition. In fact, most objections against the implications of Benign Addition in the above sequence are based on the fact that the benign additions at issue don't meet some of these extra conditions. Consider, for example, Feldman's (1995) argument that the addition of people with lives that are barely worth living make a population worse because these people are worse off than they deserve qua being people. While this argument allows Feldman to break the sequence to the Repugnant Conclusion, which requires the addition of people with lives barely worth living, it does not conflict with Superb Addition which only holds that benign additions are good if the added people have lives that are *well* worth living.

Many other objections don't apply against Superb Addition because of the condition that the additional people need to be at least as well off as the people in the original population. Consider the following objections against the implication of Benign Addition that A^+ is better than A , which are well known from Parfit's (1984, 420-25; 1986) discussion. One objection is that the average wellbeing in A^+ is lower than in A . As Carlson (1998) argues, this may well be a respect in which A^+ is worse than A even if the average view according to which average wellbeing is *all* that matters is false. Superb Addition is immune to this objection: if the additional people are at least as well off as the original people, benign additions necessarily *increase* average wellbeing, as in the transition from A to A^* .

On another objection, A^+ is not better than A because the worst-off people in A^+ are worse off than the worst-off people in A . For that reason, the maximin principle even considers A^+ *worse* than A . By contrast, if the additional people are at least as well off as

the original people, as Superb Addition requires, benign additions cannot result in populations in which the worst-off people are worse off than in the original population. For example, the worst-off people in A^* are no worse off than the worst-off people in A . This also mitigates the objection that benign additions can lead to *inequality*: with regard to equality, it seems to matter most how well off the worst-off people are. Since the worst-off people in A^+ are worse off than the people in A , it has some plausibility that the inequality in A^+ might be bad enough to make A^+ overall not better than A (although this is still hard to believe). This plausibility vanishes if benign additions lead to an unequal population in which even the worst-off people are no worse off than the people in the original population. For example, it is wholly implausible that the inequality in A^* is bad enough to make A^* not better than A .

I conclude that *if* Benign Addition is false, this is most plausibly because it does not include some of the extra conditions of Superb Addition. Hence, we should uphold Superb Addition even if the right reaction to impossibility theorems is to abandon Benign Addition. By violating Superb Addition, Neutrality pays a much higher price than impossibility theorems force us to pay.

Finally, it should be noted that the violation of Superb Addition is not *sufficient* to escape all other implausible implications of impossibility theorems either. Of course, if Superb Addition is false, Benign Addition is false too, so we cannot reach the Repugnant Conclusion via Benign Addition. However, as Arrhenius (e.g. 2000b) shows, Benign Addition can be replaced by other plausible principles, such as the combination of Non-Sadism and Non-Extreme Priority.³³ Abandoning Superb Addition therefore only helps if

³³ Roughly, Non-Sadism states that adding people with positive welfare is always better than adding people with negative welfare if the welfare of the original people is unaffected. Non-Extreme Priority reflects the idea that welfare increases for the worst-off people don't have absolute priority over welfare increases for others. See Francis (2022, 73) who uses especially compelling versions of these principles to argue against Neutrality.

we also abandon some of these other principles. Moreover, there is still the *Weak Repugnant Conclusion* according to which there is for each population of people with great lives a population of people with lives that are barely worth living that is *not worse*. As Ng (1989) and Arrhenius (2000a) show, we only need Non-Anti-Egalitarianism and the premise that benign additions are *not bad* to reach this conclusion. Neutrality agrees that benign additions are not bad.³⁴ This means that on top of violating Superb Addition (and hence Benign Addition), Neutrality also needs to violate Non-Anti-Egalitarianism or accept the Weak Repugnant Conclusion. So does the violation of Superb Addition help *at all* to avoid some other implausible implications? It only helps by allowing Neutrality to hold on to Non-Anti-Egalitarianism while accepting the Weak Repugnant Conclusion instead of the Repugnant Conclusion. But this doesn't seem like a great win: either way we are wrong about how populations A and Z compare. It certainly is not worth violating Superb Addition – especially since violating Benign Addition would suffice for this.

6. Conclusion

It has been over 20 years since John Broome demonstrated that Neutrality is greedy: if adding people to a population is neutral, it is often also neutral to add these people while making others better or worse off. According to Broome, this greediness phenomenon is responsible for several problems for which Neutrality should be rejected: the dominance problem, the conceptual problem, and the dependence problem. But many others have not been convinced and have come to accept Neutrality despite its greediness. This paper has argued that the greediness problems that Broome discusses indeed don't provide strong reason to reject Neutrality. But there is another problem with Neutrality's greediness that

³⁴ As discussed earlier, it is sufficient for the Procreation Asymmetry that mere additions of people are *not good*. Hence mere additions could be bad, which would open up the possibility that even *benign* additions can be bad (see Frick 2022). This does allow avoiding the Weak Repugnant Conclusion.

Broome discusses as an *instance* of the dependence problem: even immense welfare increases for many people would turn out neutral if they are accompanied by the addition of people with good lives. I have argued that some versions of this *welfare problem* are intuitively unacceptable and that the defences against Broome's greediness problems don't mitigate it. Even some of the implausible implications that impossibility theorems force on every population axiology pale in comparison. It is time to abandon Neutrality – and with it the Procreation Asymmetry – for good; “however strong the intuition may be, we have to grit our teeth and give it up” (Broome 2005, 410-11).

References

- Aaron, Jonas H. 2023. "Reply to Spears's 'The Asymmetry of Population Ethics.'" *Economics and Philosophy* 39 (3): 507–513.
- Adelstein, Matthew. 2024. "The Sequence Argument Against the Procreation Asymmetry." *Utilitas* 36 (4): 338–351.
- Arrhenius, Gustaf. 2000a. *Future Generations: A Challenge for Moral Theory*. Doctoral diss., Acta Universitatis Upsaliensis.
- . 2000b. "An Impossibility Theorem for Welfarist Axiologies." *Economics and Philosophy* 16: 247–266.
- . 2003. "The Very Repugnant Conclusion." In *Logic, Law, Morality: Thirteen Essays in Practical Philosophy in Honour of Lennart Åqvist*, edited by Krister Segerberg and Rysiek Śliwiński. Uppsala: Uppsala University.
- . 2022. "Population Ethics and Conflict-of-Value Imprecision." In *Ethics and Existence: The Legacy of Derek Parfit*, edited by Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan. Oxford: Oxford University Press.
- Bader, Ralf M. 2022. "The asymmetry." In *Ethics and Existence: The Legacy of Derek Parfit*, edited by Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan. Oxford: Oxford University Press.
- Benatar, David. 2006. *Better Never to Have Been: The Harm of Coming into Existence*. Oxford: Oxford University Press.

- Broome, John. 1997. "Is Incommensurability Vagueness?" In *Incommensurability, Incomparability, and Practical Reason*, edited by Ruth Chang. Cambridge, MA: Harvard University Press.
- . 2004. *Weighing Lives*. Oxford: Oxford University Press.
- . 2005. "Should We Value Population?" *Journal of Political Philosophy* 13 (4): 399–413.
- . 2007. "Replies." *Economics and Philosophy* 23: 115–124.
- . 2009. "Reply to Rabinowicz." *Philosophical Issues* 19: 412–417.
- . 2022. "Climate Change and Population Ethics." In *The Oxford Handbook of Population Ethics*, edited by Gustaf Arrhenius et al. <https://doi.org/10.1093/oxfordhb/9780190907686.001.0001>.
- Carlson, Erik. 1998. "Mere Addition and Two Trilemmas of Population Ethics." *Economics and Philosophy* 14 (2): 283–306.
- Caviola, Lucius, David Althaus, Andreas L. Mogensen, and Geoffrey P. Goodwin. 2022. "Population Ethical Intuitions." *Cognition* 218: 104941.
- Chang, Ruth. 2002. "The Possibility of Parity." *Ethics* 112: 659–688.
- Chappell, Richard Y. 2017. "Rethinking the Asymmetry." *Canadian Journal of Philosophy* 47 (2-3): 167–177.
- Dorr, Cian, Jacob M. Nebel, and Jake Zuehl. 2023. "The case for comparability." *Noûs* 57 (2): 414–453.
- Feldman, Fred. 1995. "Justice, Desert, and the Repugnant Conclusion." *Utilitas* 7: 189–206.
- Francis, Tomi. 2022. *Totalism*. DPhil Thesis, University of Oxford.
- Frick, Johann. 2014. *'Making People Happy, Not Making Happy People': A Defense of the Asymmetry Intuition in Population Ethics*. PhD diss., Harvard University.
- . 2017. "On the Survival of Humanity." *Canadian Journal of Philosophy* 47 (2-3): 344–367. <http://dx.doi.org/10.1080/00455091.2017.1301764>.
- . 2022. "Context-Dependent Betterness and the Mere Addition Paradox." In *Ethics and Existence: The Legacy of Derek Parfit*, edited by Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan. Oxford: Oxford University Press.
- Gustafsson, Johan E. 2020. "Population Axiology and the Possibility of a Fourth Category of Absolute Value." *Economics and Philosophy* 36 (1): 81–110.

- Hedden, Brian. 2024. "Parity and pareto." *Philosophy and Phenomenological Research* 109 (2): 575–592.
- Hedden, Brian, and Daniel Muñoz. 2024. "Dimensions of Value." *Noûs* 58 (2): 291–305.
- Huemer, Michael. 2008. "In Defence of Repugnance." *Mind* 117: 899–933.
- McMahan, Jeff. 1981. "Problems of Population Theory." *Ethics* 92 (1): 96–127.
- . 2013. "Causing People to Exist and saving people's lives." *The Journal of Ethics* 17 (1): 5–35.
- Mogensen, Andreas. 2023. "Philosophical Considerations Relevant to Valuing Continued Human Survival: Conceptual Analysis, Population Axiology, and Decision Theory." *Global Priorities Institute Working Paper* No. 9-2023.
- Narveson, Jan. 1973. "Moral Problems of Population." *The Monist* 57 (1): 62–86.
- Nebel, Jacob M. 2019. "Asymmetries in the Value of Existence." *Philosophical Perspectives* 33: 126–145.
- Ng, Yew-Kwang. 1989. "What Should We Do About Future Generations? Impossibility of Parfit's Theory X." *Economics and Philosophy* 5 (2): 235–253.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- . 1986. "Overpopulation and the Quality of Life." In *Applied Ethics*, edited by Peter Singer, 145–164. Oxford: Oxford University Press.
- Qizilbash, Mozaffar. 2007. "The Mere Addition Paradox, Parity and Vagueness." *Philosophy and Phenomenological Research* 75 (1).
- Rabinowicz, Wlodek. 2009. "Broome and the Intuition of Neutrality." *Philosophical Issues* 19: 389–411.
- . 2022. "Getting Personal: The Intuition of Neutrality Reinterpreted." In *The Oxford Handbook of Population Ethics*, edited by Gustaf Arrhenius et al. <https://doi.org/10.1093/oxfordhb/9780190907686.001.0001>.
- Roberts, Melinda. 2011. "The Asymmetry: A Solution." *Theoria* 77 (4): 333–367.
- . 2022. "The Person-Based Intuition and the Better Chance Puzzle." In *The Oxford Handbook of Population Ethics*, edited by Gustaf Arrhenius et al. <https://doi.org/10.1093/oxfordhb/9780190907686.001.0001>.
- Spears, Dean. 2019. "The Asymmetry of Population Ethics: Experimental Social Choice and Dual-Process Moral Reasoning." *Economics and Philosophy*: 1–20.
- Steele, Katie. 2025. "Longtermism and Neutrality About More Lives." In *Essays on Longtermism*, edited by James Barrett, Hilary Greaves, and David Thorstad. Oxford: Oxford University Press.

- Tännsjö, Torbjörn. 2002. "Why We Ought to Accept the Repugnant Conclusion." *Utilitas* 14: 339–359.
- Tarsney, Christian, Harvey Lederman, and Dean Spears. 2025. "A Dominance Argument Against Incompleteness." *Philosophical Review* 134 (4): 455–490.
- Thomas, Teruji. 2018. "Some Possibilities in Population Axiology." *Mind* 127 (507): 807–832.
- Thornley, Elliott. 2022. "Critical Levels, Critical Ranges, and Imprecise Exchange Rates in Population Axiology." *Journal of Ethics & Social Philosophy* 22: 382.
- Wilkinson, Hayden. 2023. "How to Neglect the Long Term." *Global Priorities Institute Working Paper* No. 11-2023.

Chapter 2: In Defence of Pseudo-Dominance

Abstract

According to pseudo-dominance principles, something is overall better if it is better in some respect and not worse in any respect. Normally, these principles are relativized to certain sets of evaluatively relevant respects, such as individual wellbeing levels or value dimensions, and they are rejected because of problems arising from incommensurateness. This paper argues that these incommensurateness problems can be avoided if the pseudo-dominance relation is not relativized to any particular set of respects. But this comes at the cost of making the pseudo-dominance principle inapplicable. In response to this application problem, two refined pseudo-dominance principles are developed that are both applicable and avoid the problems from incommensurateness. The final part of the paper applies these refined pseudo-dominance principles to population ethics. It argues that they refute the view that the existence of additional happy people is a neutral matter.

1. Introduction

Imagine that you need to compare the value of two things, which could be two different vaccines, climate policies, or any two possible states of the world. There are many respects in which the two things differ and several of these respects are relevant for their evaluation. It is not immediately obvious, but after some research and deliberation it turns out that there are some respects in which one thing is better than the other, but that there is not a single respect in which the other thing is better. You conclude, therefore, that the first thing is better overall. That seems reasonable. Indeed, you would make the same type of inference

that numerous philosophers have made. Consider the following examples that compare the value of different populations (i.e. sets of people with assigned wellbeing levels): “I conclude that, since Alpha is in this way better than A+, and is in no way worse, Alpha is better than A+” (Parfit 1986, 158-59). “So *C* is worse than *A* in one respect, and neither better nor worse in the other respect. Intuitively, therefore, *C* must on balance be worse than *A*” (Broome 2005, 409). “If we accept the Intuition of Neutrality, we therefore think that A+ is in no respect better than A. Hence, if in the 3-possible case the presence of injustice makes A+ worse than A in *some* respect, this is enough to make A+ worse than A *all things considered*” (Frick 2022, 253).

When making these inferences, Parfit, Broome, and Frick appear to rely on the principle that if one of two things is better in some respect and not worse in any respect, it is better overall. *Prima facie*, this principle is extremely plausible, and it is natural to make inferences according to it. However, many variants of the principle that are relativized to certain sets of evaluatively relevant respects have come under attack in the recent literature. Among them are the principle that something is better overall if it is better with respect to some *individual’s wellbeing* and not worse with respect to any individual’s wellbeing, the principle that something is (ex ante) better overall if it is better with respect to the *value in one possible state of the world* and not worse with respect to the value in any possible state of the world, and the principle that something is better overall if it is better with respect to some *value dimension* (e.g. pleasure, equality, or beauty) and not worse with respect to any value dimension.³⁵ All these principles have been rejected on the basis that they propose sufficiency conditions for overall betterness that only require that the supposedly better

³⁵ One may wonder if the latter principle that refers to value dimensions is really more specific than the general principle that refers to respects, or whether it is just a restatement of that principle. Some authors seem to use both terms roughly interchangeably (e.g. Nebel 2019, 139). But, as I will explain in the next section, other authors (e.g. Hedden & Muñoz 2024) use “value dimension” in a more narrow sense in which value dimensions correspond to (what value theories acknowledge as) fundamental values, as indicated by the examples of pleasure, equality, and beauty.

thing is *not worse* than the other thing in certain respects rather than requiring that it is *at least as good* as the other thing in these respects (e.g. Hare 2013, Nebel 2019, Hedden 2024, Hedden & Muñoz 2024). This has been taken to make a crucial difference because being not worse in a respect – unlike being at least as good in a respect – is compatible with being *incommensurate* in that respect, i.e. neither better nor worse but also not equally as good.³⁶ For that reason, the above principles have been described as *pseudo-dominance* principles (Nebel 2019, 139): they superficially resemble dominance principles but fail to qualify as such because their antecedents don't rule out that the compared things are incommensurate in some respect. But then, it is argued, the pseudo-dominant thing could also be incommensurate with the pseudo-dominated thing *overall*; it does not need to be better overall, as the pseudo-dominance principles claim.

If all of this is right and we need to reject the various principles that relativize the pseudo-dominance relation (i.e. the relation of being better in some respect and not being worse in any respect) to certain sets of respects, what does this mean for the simple non-relativized pseudo-dominance principle, according to which something is better overall if it is better in some respect and not worse in any respect? This question is rarely discussed.³⁷ Presumably, that is because it seems *obvious* that the non-relativized principle needs to be rejected along with its relativized versions.³⁸ After all, it is also a pseudo-dominance principle that employs the relation *not worse than*. If this relation is compatible with

³⁶ This stipulative definition of “incommensurateness” will be used throughout the article.

³⁷ As commented in footnote 1, some authors seem to use “dimension” interchangeably with “respect”. If this is right, these authors could be interpreted as arguing against the non-relativized pseudo-dominance principle. See also Parfit (2016, 115) who argues that “the relation *not worse than* does not imply *at least as good as*, so one of two things can be in one way better and in other ways not worse without being better all things considered.” In general, it is difficult to tell whether an author argues against the non-relativized principle since there is a relativized principle, which I will introduce in the next section as “Strong Pseudo-Dominance”, that is easily confused with it.

³⁸ One might even think that the more specific principles that are relativized to certain sets of respects are *entailed* by the simple pseudo-dominance principle, given that the sets of respects to which they are relativized fully determine overall value. As I will explain in the next two sections, this is not the case given a strict interpretation of the simple principle. Instead, they are entailed by Strong Pseudo-Dominance.

incommensurateness, and the relativized principles need to be rejected *because* their antecedents don't rule out incommensurateness, it seems that the non-relativized principle needs to be rejected too. Although seldomly made explicit, this lesson seems to be commonly drawn from the discussion of the relativized pseudo-dominance principles.

This paper argues that there is good reason to uphold the simple pseudo-dominance principle even if the arguments against the relativized principles are sound. If we take seriously its condition that the pseudo-dominant thing is not worse in *any* respect, we obtain a principled answer to the incommensurateness problem. Roughly, if there is not a single respect in which the pseudo-dominant thing is worse, then there is also no *value conflict* between the compared things, and if there is no value conflict between them, they plausibly cannot be incommensurate either. Since the relativized principles only require that the pseudo-dominant thing is not worse in any respect of a certain *kind*, such as respects that correspond to value dimensions or wellbeing levels, this response to the incommensurateness problem is not available to the relativized principles. We therefore have good reason to uphold the simple pseudo-dominance principle even if we reject its relativized variants due to incommensurateness.

However, there is a distinct problem for the simple pseudo-dominance principle: its condition that there is not a single respect in which the pseudo-dominant thing is worse makes the principle more plausible, but also less applicable. It is easy to see that this condition is barely ever met in comparisons between actual things – we can virtually always find at least some minor respect in which each of two things is better than the other. But I will argue that there is also a more fundamental application problem: the principle is not even applicable in hypothetical cases because, at least normally, betterness in some respect *necessitates* worseness in another respect. In response to this application problem, I will discuss several possible refinements of the simple principle that allow the pseudo-dominant thing to be worse in certain respects: *gerrymandered* respects, *evaluatively redundant*

respects, and *outweighed* respects. I will argue that the latter two options lead to pseudo-dominance principles that are both applicable and defensible – in particular against objections from incommensurateness.

Finally, I will put the two refined pseudo-dominance principles to work. I will show how they can be used to argue against the “neutrality intuition” in population ethics, according to which the existence of additional people with good lives does not make the world better. The argument is analogous to Broome’s “greediness argument”, which employs the principle that “a good thing plus a neutral thing should add up to a good thing” (2005, 409). Broome’s argument has been rejected because this principle seems incompatible with incommensurateness (Frick 2017, 355; Hedden 2024, 579). I will show that there is an amended version of this argument that employs my refined pseudo-dominance principles instead and thus avoids the problems with incommensurateness.

Here is the structure of the paper in more detail: Section 2 discusses under which conditions something is better in a respect and distinguishes on that basis the simple pseudo-dominance principle, which I will call *Weak Pseudo-Dominance*, from a logically stronger principle, *Strong Pseudo-Dominance*, which relativizes the pseudo-dominance relation to sets of respects on which overall betterness supervenes. Section 3 discusses common objections from incommensurateness against various pseudo-dominance principles. It argues that these targeted principles are best understood as instances of Strong Pseudo-Dominance and that the objections refute that principle while Weak Pseudo-Dominance is immune to them. The reason is that all these objections are based on cases of incommensurateness in which each of the compared things is better in some respect, so that Weak Pseudo-Dominance does not apply. Section 4 argues more generally that Weak Pseudo-Dominance is compatible with all compelling cases of incommensurateness because all such cases, at least within a certain domain, require some sort of value conflict. It argues that if this requirement holds, Weak Pseudo-Dominance is most plausibly true.

Section 5 introduces the application problem, according to which Weak Pseudo-Dominance is not even applicable in hypothetical cases. Section 6 discusses different possible solutions to the application problem and argues for two logically stronger versions of Weak Pseudo-Dominance: *Moderate Pseudo-Dominance* and *Generalized Pseudo-Dominance*. Section 7 applies these principles to argue against the neutrality intuition. Section 8 summarizes.

2. Weak and Strong Pseudo-Dominance

We are used to comparing things in all kinds of respects. Some ways of travelling are better with respect to their environmental impact while other ways of travelling are better with respect to comfort. Some diets are better with respect to health while others are better with respect to gustatory pleasure. Some climate policies are better with respect to the welfare of present people while others are better with respect to the welfare of future people. And so on.

Philosophers sometimes say that things are better or worse along a *value dimension*. By this, they typically mean something more specific than just a respect in which things can be better or worse; it is supposed to be a more demanding notion. For example, Hedden & Muñoz (2024, 291) introduce the notion of a value dimension by saying that only value pluralists believe in several dimensions of value while value monists, such as hedonists, hold that there is only one dimension of value. In that narrow sense, in which value dimensions correspond to (what is intuitively thought of as) fundamental values such as pleasure or knowledge, there are many respects in which things can be better or worse that are not value dimensions.³⁹ After all, even value monists can compare things in various

³⁹ Later on, Hedden & Muñoz suggest the following definition: “a dimension is (i) a genuine and distinct respect in which things can be good or bad, (ii) to which overall value is responsive” (2024, 297). I take it that overall value is responsive to any (genuine) respect in which things can be good or bad, so what makes dimensions a more narrow category is presumably their “distinctness”.

respects. The examples I gave above are a case in point. They are not tied to any particular value theory; any minimally plausible theory is compatible with the judgement that one way of travelling is better with respect to its environmental impact while the other is better with respect to comfort, for example. To make this judgement, one doesn't have to believe that the environment and comfort are two fundamental values. Their value may derive from other values, which are fundamental, or there may be no such distinction in the first place. The notion of being better in a respect does not require the notion of fundamental values or value dimensions – it doesn't require that there is a hierarchy between values.

We can think of the respects in which things can be better or worse as the factors that jointly determine overall betterness. Overall betterness *supervenes* on certain sets of respects: any two things that are exactly alike in each respect in these sets are also overall better than exactly the same things; if X but not Y is better than some third thing, X and Y must differ in at least one respect in these sets.⁴⁰ Crucially, there are *many* such “supervenience sets” that don't, or only partially, overlap; not every respect is included in each supervenience set.⁴¹ Which sets of respects are supervenience sets is of course not a trivial question; different axiologies imply different answers to it. But according to every axiology, there are many supervenience sets.

⁴⁰ Does this mean that X and Y are *equally good* if they are alike in all respects of such a set on which overall betterness supervenes? No, but it follows given minimal auxiliary assumptions. Given that nothing is better than itself, X cannot be better or worse than Y if X and Y are better than exactly the same things. This leaves two possibilities: they could be equally good or incommensurate. If they were incommensurate, then, given that incommensurateness is insensitive to mild sweetenings and sourings, X should also be incommensurate with some slightly worsened version of Y (or Y should be incommensurate with a slightly worsened version of X). This is impossible if X and Y are better than exactly the same things since, of course, Y is better than its slightly worse version. Thus, X and Y are equally good if they are alike in all respects of a set on which overall betterness supervenes.

⁴¹ Not every supervenience base of overall betterness is a set of respects in which things can be better or worse. For example, overall betterness also supervenes on the set {overall worseness, overall equal goodness, overall incommensurateness}. The properties in this set are not respects, or at least they aren't respects in which things can be better or worse. The set therefore doesn't count as a “supervenience set” in my sense.

Consider the total (utilitarian) view according to which X is better than Y if and only if X has a higher total wellbeing than Y. Obviously, overall betterness supervenes on the respect of total wellbeing, on that view, and hence on the set {total wellbeing} that only contains this respect. But on the same view, overall betterness also supervenes on {total human wellbeing, total non-human wellbeing}, on {total shrimp wellbeing, total non-shrimp wellbeing}, and on many other sets of respects. That is because {total wellbeing} supervenes on each of these sets. Due to the transitivity of the supervenience relation, this means that overall betterness supervenes on each of these sets as well: overall betterness (on the total view) supervenes on total wellbeing, which in turn supervenes on e.g. {total human wellbeing, total non-human wellbeing}. Thus, by transitivity, overall betterness supervenes on {total human wellbeing, total non-human wellbeing}. The same holds for many other sets of respects, such as the set of all individual wellbeing levels, the set of all welfare goods, and so on.

The total view can even be represented by *value functions* that depend on the respects in these sets.⁴² For example, each of the following value functions represents the total view.

$$V_1(X) = (\text{total wellbeing in } X)$$

$$V_2(X) = (\text{total human wellbeing in } X) + (\text{total non-human wellbeing in } X)$$

$$V_3(X) = (\text{total shrimp wellbeing in } X) + (\text{total non-shrimp wellbeing in } X)$$

$V_1(X)$ obviously represents the total view. Since, for any X, $V_1(X) = V_2(X) = V_3(X)$, $V_2(X)$ and $V_3(X)$ represent the total view as well. These functions tell us *how* the respects in each

⁴² A value function $V(X)$ represents a betterness ordering, such as the total view, if and only if, for all X and Y, $V(X) > V(Y)$ if and only if X is better than Y according to that betterness ordering. Since not all betterness orderings can be represented by a value function (given that the range of the value function is the real numbers), it is not a general requirement for a supervenience set that there is a corresponding value function. But it does hold conversely that if a betterness ordering can be represented by a value function that only depends on the respects in a certain set, then this set is a supervenience base of overall betterness according to this ordering.

supervenience set determine overall betterness and, accordingly, whether things are *better* in a respect if their (descriptive) value in that respect is higher. Here, all three functions strictly increase with the values of each respect of one supervenience set. For example, $V_2(X)$ increases with both total shrimp wellbeing in X and total non-shrimp wellbeing in X . This means that, as long as we hold all other respects of the respective supervenience set fixed, increasing the value of one respect of that set necessarily leads to something better overall. For example, increasing total shrimp wellbeing while holding total non-shrimp wellbeing fixed necessarily leads to something better overall. This shows that if X has a higher total shrimp wellbeing than Y , this is a respect in which X is better than Y : it makes X overall better than Y if all else in the supervenience set is equal between X and Y .⁴³

I have inferred from the overall betterness ordering of the total view that, on that view, X is better than Y with respect to total shrimp wellbeing if X has a higher total shrimp wellbeing than Y . Intuitively, this is the right result. For illustration, imagine that you compare two different environmental policies based on the total view. One way to do this would be to compare the policies based on their impacts on the total wellbeing of each species. After going through all species and summing up their total wellbeings for each policy, it turns out that policy X would lead overall to a higher total wellbeing than policy Y . Clearly, this makes policy X overall better on the total view.⁴⁴ Still, if this policy would lead to a lower total wellbeing for some species, such as (a certain species of) shrimp, this seems to be a respect in which this policy is worse. That it is worse in this respect could motivate the search for a third policy that combines the advantages of the other two policies:

⁴³ It is tempting to *analyse* the concept of being better in a respect in terms of this connection with overall betterness: X is better than Y with respect to R if and only if there is a supervenience set S including R such that, for any two things A and B that (descriptively) compare with respect to R in the same way in which X and Y compare with respect to R , if A and B are exactly alike in all other respects in S , A is overall better than B . This analysis seems plausible to me, but I won't rely on it. What this analysis holds to be a sufficient and necessary condition for betterness in a respect may only be a sufficient condition.

⁴⁴ Strictly speaking, what is better on the total view is not the policy but the policy's *outcome* since the policy itself does not have a total wellbeing.

a policy that is as good as Y with respect to total shrimp wellbeing while being as good as X with respect to the total wellbeing of other species.

Let's consider a more complex, non-welfarist axiology according to which overall betterness supervenes on {total wellbeing, knowledge, beauty}. The respects in this set could be the fundamental values, or value dimensions, according to this axiology, so that this particular supervenience set is also the set of all value dimensions. But again, there will be many alternative supervenience sets. We can identify them by replacing some or all respects in the above set by (sets of) respects on which the former supervene. For example, we could replace the respect of total wellbeing by any set of respects on which total wellbeing supervenes, such as by {total human wellbeing, total non-human wellbeing}. The resulting set would be {total human wellbeing, total non-human wellbeing, knowledge, beauty}, which is another supervenience set given the non-welfarist axiology. Thus, for each alternative supervenience set that exists on the total view, there will be a corresponding alternative supervenience set on this more complex view. Additionally, there are further supervenience sets on the non-welfarist view that result from replacing the other respects in the "original" supervenience set by subvening sets of respects. For example, {knowledge} could be replaced by {scientific knowledge, non-scientific knowledge} and {beauty} could be replaced by {natural beauty, non-natural beauty}. The combination of all these replacements results in the following supervenience set that does not overlap at all with the original set: {total human wellbeing, total non-human wellbeing, scientific knowledge, non-scientific knowledge, natural beauty, non-natural beauty}. I have not specified the betterness ordering of the non-welfarist view, so we cannot infer in which case the view will say that X is *better* than Y in any of these respects. But it is clear what a proponent of the view would say about this, and what a plausible specification of the view will imply: X is better than Y in any of these respects if X has *more* of them than Y, e.g. if

X contains more total human wellbeing, more non-scientific knowledge, and more natural beauty.⁴⁵

Finally, consider an axiology on which overall betterness supervenes on the set of total wellbeing and some measure of inequality. Depending on the particular measure of inequality, this could be an egalitarian or a prioritarian axiology.⁴⁶ Whatever measure of inequality is chosen, as long as it is a measure of the inequality between the wellbeing levels of individuals (rather than, e.g., inequality of *opportunity* for wellbeing), inequality supervenes on the set of all individual wellbeing levels. Of course, total wellbeing supervenes on that set as well. Hence, on some egalitarian and prioritarian views, overall betterness supervenes both on {total wellbeing, inequality} and on {individual 1's wellbeing, individual 2's wellbeing, ... individual n's wellbeing}, where n is the number of people who exist in X or Y. Again, we have two non-overlapping supervenience bases of overall betterness. There are well-known egalitarian and prioritarian value functions that depend on the respects in these sets, such as the prioritarian square-root function (i.e. the sum of the square roots of each individual's wellbeing). But even without looking at these particular functions, it is intuitively clear in which case plausible prioritarian and egalitarian views will say that something is better in the respects of the two supervenience sets: X is better than Y with respect to total wellbeing if X has higher total wellbeing than Y, X is better than Y with respect to inequality if X has less inequality than Y, and X is better with

⁴⁵ At least, this holds for any plausible *measure* of these non-welfarist values. This measure does not have to be the *total* of these values; it could also be the average or something more complicated. But whatever the correct measure is, it seems plausible that X is better than Y with respect to the non-welfarist value if X has more than Y of whatever is the adequate measure of that value.

⁴⁶ See Fleurbaey (2015, 207) and Broome (2019, 593) who argue that (not only egalitarianism but also) prioritarianism can be represented by a value function that depends on total wellbeing and a measure of inequality.

respect to some individual's wellbeing if that individual's wellbeing is higher in X than in Y.⁴⁷

I have argued that, independently of which axiology we assume, there are many different sets of respects on which overall betterness supervenes. This is important because it shows that the following two principles are not equivalent.

Weak Pseudo-Dominance: For all X and Y, if X is better than Y in some respect and X is not worse than Y in any respect, then X is better than Y overall.

Strong Pseudo-Dominance: For all X and Y, if there is a supervenience set that contains a respect in which X is better than Y and that does not contain any respect in which X is worse than Y, then X is better than Y overall.

Weak Pseudo-Dominance is the simple pseudo-dominance principle that I seek to defend. If overall betterness supervened only on the total set of *all* respects in which things can be better or worse, Weak Pseudo-Dominance would be equivalent with Strong Pseudo-Dominance. But since there are many disjoint supervenience sets, Weak Pseudo-Dominance could be true even if Strong Pseudo-Dominance is false. According to the latter, we only need to find one supervenience set in which X pseudo-dominates Y to conclude that X is overall better than Y. The sufficiency condition for overall betterness that Weak Pseudo-Dominance posits, by contrast, requires that there is no respect whatsoever in which X is worse than Y. Accordingly, there may not be any respect in which X is worse than Y outside of any one supervenience set either. In particular, X may not be worse than Y in

⁴⁷ There is an (implausibly) extreme form of egalitarianism according to which increasing someone's wellbeing while holding everyone else's wellbeing fixed sometimes leads to an overall worse outcome, i.e. a version in which "levelling-down" is sometimes overall good. On this view, it is less clear whether we should say that X is better than Y with respect to some individual's wellbeing if her wellbeing is higher in X than in Y. But plausibly, this is still the right thing to say: in these cases, X is worse than Y *despite* being better with respect to the person's wellbeing and *because* X is worse than Y with respect to equality.

any respect in any of the *other* supervenience sets. If all respects in which things can better or worse are part of *some* supervenience set, Weak Pseudo-Dominance is equivalent with the following principle.

Weak Pseudo-Dominance*: For all X and Y, if there is a supervenience set that contains a respect in which X is better than Y and there is no supervenience set that contains any respect in which X is worse than Y, then X is better than Y overall.

In fact, this principle is *trivially* equivalent with Weak Pseudo-Dominance since each respect in which things can be better or worse is part of at least one supervenience set: the total set of all respects in which things can be better or worse.⁴⁸ Still, this equivalent version is useful in making explicit what satisfying the sufficiency condition of Weak Pseudo-Dominance requires on top of satisfying the sufficiency condition of Strong Pseudo-Dominance: instead of merely requiring that X is not worse than Y in any respect of *one* supervenience set, it requires that X is not worse than Y in any respect of *any* supervenience set. As we have seen, there are many supervenience sets even on relatively simple axiologies like the total view. For Weak Pseudo-Dominance to apply, it would need to be the case that X is not worse than Y in any of the respects in any of these sets. So, for example, X would need to be not worse than Y with respect to the total wellbeing of each species, with respect to, even, the wellbeing of each individual, with respect to the total amounts of each welfare good, such as total pleasure, and so on. This means that Weak Pseudo-Dominance implies overall betterness in far fewer cases than Strong Pseudo-Dominance. For the latter, it would suffice, given the total view, that X is better than Y

⁴⁸ The non-trivial question is whether each respect in which things can be better or worse is part of a *minimal* supervenience set, i.e. a supervenience set that is not a superset of a supervenience set and some other (non-empty) set of respects. Again, I think that the answer is plausibly 'yes', but I won't argue for it here. (But consider that all the examples of supervenience sets that I gave above are minimal supervenience sets.) For my purpose, it would be fine if X could be worse than Y in some respect even if no minimal supervenience set contains a respect in which X is worse than Y.

with respect to total wellbeing, or alternatively that X is better than Y with respect to total human wellbeing and not worse than Y with respect to total non-human wellbeing.

So the sufficiency condition of Weak Pseudo-Dominance is far more demanding than the sufficiency condition of Strong Pseudo-Dominance. Does this difference affect the relative plausibility of the two principles? Intuitively, it does. *Prima facie*, it seems plausible that something must be better if it is better in some respect and not worse in any respect within some supervenience set. But this becomes at least somewhat doubtful once we realize that there could still be worseness in plenty of other respects. It seems safer to only hold that there must be overall betterness if there no worseness in any respect whatsoever. Assume that overall betterness supervenes on {total wellbeing, knowledge, beauty} and that X is better than Y with respect to knowledge and not worse than Y with respect to total wellbeing and beauty. *Prima facie*, this suffices to make X overall better than Y. But now assume additionally that X is worse than Y with respect to the wellbeing of many individual people, with respect to certain welfare goods, such as the enjoyment of music, and with respect to the beauty of certain landscapes. Intuitively, it becomes less clear whether X is overall better than Y. By contrast, if we assume instead that X is not worse than Y in any of these respects either, it seems *obvious* that X is overall better than Y. If X and Y were options between which we can choose, the choice would be an easy one to make.

3. Objections from Incommensurateness

I have argued that Weak Pseudo-Dominance is significantly weaker and intuitively more compelling than Strong Pseudo-Dominance. In this section, I will argue that Weak Pseudo-Dominance is also immune to the objections from incommensurateness that have been made against various pseudo-dominance principles. These principles are closely related to

Strong Pseudo-Dominance and the objections show that, conditional on the existence of incommensurateness, Strong Pseudo-Dominance needs to be rejected along with most of these related principles. (There is one related pseudo-dominance principle that may be defensible against the objections.) But Weak Pseudo-Dominance is unaffected since all objections assume – sometimes explicitly and sometimes implicitly – that there are some respects in which the pseudo-dominant thing is worse.

3.1 Paradigmatic Cases of Incommensurateness

Various pseudo-dominance principles have been rejected on the grounds that they are incompatible with paradigmatic cases of incommensurateness. The supposed reason for this incompatibility is, roughly, that these pseudo-dominance principles rule out that incommensurateness can persist after slightly improving (or “sweetening”) one of the incommensurate things in a respect in which they were originally equally good. The such improved thing will then be better in this respect and incommensurate – and hence not worse – with respect to the remaining difference. Given certain specifications of the involved respects, various pseudo-dominance imply that the improved thing is better overall. But paradigmatic cases of incommensurateness suggest that this doesn’t need to be the case – there could still be incommensurateness after the improvement.

Frick (2017, 355) argues along these lines against the *decomposition principle*, which roughly states that X is overall better than Y if the total difference between X and Y can be decomposed into two differences such that X is better than Y with respect to one difference and X is not worse than Y with respect to the other difference.⁴⁹ To argue against

⁴⁹ Frick formulates the principle in terms of changes rather than differences, and he specifies that one of the changes must be between incommensurate things: “if an overall change, such as the move from A to C, can be decomposed into (i) a change for the worse – such as the move from A to C1 – plus (ii) a move between two things that are incommensurate in value – such as the move from C1 to C –, it must overall be a change

this principle, Frick uses the common example of two incommensurate careers (cf. Raz 1986, 332). Concretely, a career as a musician, M, could be incommensurate both with a career as a painter, P, and with an improved version of the latter career with a higher salary, P+. Assume that M and P have the same salary, so that P+ has a higher salary than both P and M.⁵⁰ Then, the total difference between P+ and M can be decomposed into the difference between P+ and P (i.e. the difference in salary) and the difference between P and M (i.e. all the other differences between the two careers). With respect to the former difference, P+ is better than M, and with respect to the latter difference P+ is not worse than (but incommensurate with) M. The decomposition principle therefore implies that P+ must be better than M; it cannot account for the incommensurateness of the two careers.

The decomposition principle is closely related to Strong Pseudo-Dominance. Both state that X is overall better than Y if X pseudo-dominates Y within a certain *set* of respects. In case of Strong Pseudo-Dominance, this is a set on which overall betterness supervenes. In case of the decomposition principle, this is a set of respects that decomposes the total difference between X and Y. Both these sets have in common that, even though the respects that they contain fully determine overall betterness between X and Y,⁵¹ they don't include many respects in which X could be better or worse than Y. I have already shown this for Strong Pseudo-Dominance. In case of the decomposition principle, the reason is slightly different: a set that decomposes the complete difference between X and Y into two

for the worse" (2017, 355). I take it that these differences are superficial and that my formulation makes the principle neither more nor less plausible.

⁵⁰ Without this assumption it wouldn't be necessarily the case that the difference between P+ and M can be decomposed into the difference between P+ and P and the difference between P and M. Assume, for example, that P+ and M have the same salary. Then, the difference between P+ and P isn't part of the difference between P+ and M: having a different salary is not part of the difference between two careers with the same salary.

⁵¹ A set of respects that decomposes the total difference between X and Y is itself a kind of supervenience set: it is a supervenience set for the limited class of things that are alike in all the respects in which X and Y are alike. For any *such* two things, if they are (additionally) alike in all respects of a set that decomposes the difference between X and Y, they are overall better than the same things.

differences reflects each part of that difference, not only their difference in certain evaluatively relevant respects within a supervenience set. But since the decomposition is very coarse-grained – it decomposes the total difference into *two* differences – being not worse with respect to each part of the decomposed difference is not sufficient for being not worse in any respect: there could still be worseness with respect to more fine-grained parts of the total difference. In the case at issue, this means that the improved career as a painter P^+ could be worse than the career as a musician M in some respects, even though P^+ is neither worse than M with respect to salary nor worse than M with respect to their remaining difference, which corresponds to the total difference between M and P . In fact, there are many plausible candidates of such respects: as a musician, one could refine one's musical skills, develop a deeper understanding of music, and have more enjoyment from playing music, for example. If M is better than P^+ in any of these respects, *Weak Pseudo-Dominance* is silent on how M and P^+ compare. As far as *Weak Pseudo-Dominance* is concerned, M and P^+ could be incommensurate.

But couldn't a career as a painter be no worse than a career as a musician in any respect whatsoever? That is hard to imagine but arguably possible.⁵² If the career as a painter is that amazing that it allows one to enjoy all the advantages as a musician, or if the career as a musician is that unsuccessful (or that unfitting to one's personality) that one does not gain anything valuable from it, there might be no respect in which the career as a painter is worse. If M and P are such careers, and P^+ is an improved version of P , *Weak Pseudo-Dominance* implies that P^+ is better than M . But that seems to be exactly the right verdict. If there is nothing about M that is better than P and P^+ , it ceases to be plausible that M is incommensurate with P and P^+ . That they are incommensurate gains plausibility

⁵² As I will argue in section 5, it may be *generally* impossible for something to be better in some respect and not worse in any respect. I set this application problem aside in the discussion in this section and assume that, given sufficient stipulations, *Weak Pseudo-Dominance* can apply in (hypothetical) cases.

from the fact that there are distinct advantages of each career that are not (or not precisely) comparable. Once we assume that only one career has any distinct advantage over the other, it becomes plausible that this career is better overall.

Consider a different case that Hedden (2024, 589) uses to argue against the *dimension-wise* pseudo-dominance principle that X is overall better than Y if X is better than Y along some value dimension and not worse than Y along any value dimension. In this case, you have the choice between saving one of two valued possessions: your wedding album, W, and your first edition of Descartes' *Meditations*, M. Hedden argues that these two options could be incommensurate along the "coarse-grained dimension of sentimental-cum-cultural value" (ibid.) while being equally good with respect to the dimension of money. Accordingly, an improved version of option W that additionally gives you 5 dollars, W+, would be better than M along the dimension of money while being incommensurate with M – and hence not worse than M – along the dimension of sentimental-cum-cultural value. If the two options don't differ along any other dimension, dimension-wise pseudo-dominance implies that W+ is overall better than M. But intuitively, M and W+ could well be incommensurate.

Given *Dimensionalism* – the view that overall value relations supervene on the set of all value dimensions (see Hedden & Muñoz 2024, Hedden & Nebel 2024) – dimension-wise pseudo-dominance is an instance of Strong Pseudo-Dominance. It holds that it is sufficient for overall betterness that X pseudo-dominates Y in a particular supervenience set: the set of all value dimensions. Accordingly, Hedden's objection also affects Strong-Pseudo Dominance. In fact, his objection is *more* compelling as an argument against Strong Pseudo-Dominance than as an argument against dimension-wise pseudo-dominance. That is because it is doubtful that money and sentimental-cum-cultural value are value dimensions. They surely aren't in the sense that serves to distinguish value monists from value pluralists: money is not a fundamental value, and sentimental-cum-cultural value may

be the conjunction of two fundamental values but it is not *one* fundamental value. Hedden seems to think here just of *respects* in which things can be better or worse, and perhaps he even thinks of (something like) Strong Pseudo-Dominance as the target of his objection. Against the latter principle, his argument can be restated as follows: W^+ is better than M with respect to money, and W^+ is not worse than M with respect to sentimental-cum-cultural value. There is a supervenience set that includes these respects and otherwise only respects in which M and W^+ don't differ. Thus, W^+ pseudo-dominates M within a supervenience set. It follows by Strong Pseudo-Dominance that, against the common intuition, W^+ is overall better than M .

But clearly the argument does not affect Weak Pseudo-Dominance. In this case, it is even transparent that the pseudo-dominant thing is worse in a respect. W^+ is incommensurate with M with respect to sentimental-*cum*-cultural value as a consequence of being *better* with respect to sentimental value and *worse* with respect to cultural value. Each of W^+ and M (and each of W and M) is better than the other in some respect. Thus, Weak Pseudo-Dominance does not apply. As before, we could change the case such that W^+ is not worse than M with respect to cultural value or any other respect. We could imagine, for example, that the particular wedding album at issue also has high cultural value or that the book is just a recent copy of the first edition of Descartes' *Meditations*. Given such a stipulation, Weak Pseudo-Dominance does imply that W^+ is overall better than M . But again, this seems to be the correct verdict. What originally motivated the intuition of incommensurateness was the value conflict between sentimental value and cultural value.

Finally, here is Hedden's (2024, 590) argument against *individual-wise* pseudo-dominance, which states that X is overall better than Y if X is better than Y for some individual and not worse than Y for any individual. There are two populations that only differ with respect to the life of one individual, who is a philosopher in population P and a journalist in population J . Intuitively, P and J could be incommensurate for this one

individual. Now consider an improved version of P, P+, in which the wellbeing of another individual is minimally increased. P+ is better than J for this individual and not worse than J for any other individual. Thus, by individual-wise pseudo-dominance, P+ is overall better than J. But intuitively, they could well be incommensurate.

Given *Welfarism* – the view that overall value relations supervene on the set of all individual wellbeing levels – individual-wise pseudo-dominance is an instance of Strong Pseudo-Dominance. It holds that it is sufficient for overall betterness that X pseudo-dominates Y in a particular supervenience set: the set of all individual wellbeing levels. Given that the intuition that P+ and J are incommensurate is correct, Hedden’s argument refutes individual-wise pseudo-dominance and, given Welfarism, also Strong Pseudo-Dominance. But Weak Pseudo-Dominance is unaffected because each population is better than the other in some respect. In particular, J is better than W+ with respect to whatever welfare goods favour a career as a journalist over a career as a philosopher. If there weren’t any such welfare goods, it would cease to be plausible that J and W+ are incommensurate. W+ would be better than J, as Weak Pseudo-Dominance would then imply.

I will argue in the next section that it is no coincidence that all cases of incommensurateness that have been used to argue against pseudo-dominance principles involve some kind of value conflict: there are no compelling examples of incommensurateness without any value conflict, at least within a suitably restricted domain of evaluands. But before making this argument, it is worth noting that Weak Pseudo-Dominance can respond to other common objections against pseudo-dominance principles in a similar way: the incommensurateness that these objections assume either involves some value conflict or is not justified. I will show this for two common objections in the next two subsections.

3.2 Cyclicity

The reason pseudo-dominance principles are most commonly (and most decisively) rejected is that they lead to betterness-cycles: if these principles were true, it would sometimes be the case that something is better than a second thing, which again is better than a third thing, while the third thing is better than the first thing (e.g. Hare 2013, 176; Nebel 2019, 139; Hedden 2024, 578; Hedden & Muñoz 2024, 300). The argument that intends to show that pseudo-dominance principles imply cyclicity is based on cases with the following structure. There are three things A, B, and C and three respects R1, R2, and R3 in which they differ. Each thing has value x , x^+ , or y in each respect, where x^+ is better than x , and both x and x^+ are incommensurate with y . More specifically, the following values are assigned to A, B, and C.

	R1	R2	R3
A	x	x^+	y
B	y	x	x^+
C	x^+	y	x

Now take any pseudo-dominance principle that implies that, for a suitable specification of the respects R1, R2, and R3, if X is better than Y in one of these respects and not worse than Y in the other two respects (i.e. if X pseudo-dominates Y in $\{R1, R2, R3\}$), X is overall better than Y. For such a pseudo-dominance principle, cyclicity can be demonstrated as follows: A is better than B with respect to R2 and incommensurate with B in the other two respects. Thus, A is better than B. B is better than C with respect to R3 and incommensurate with C in the other two respects. Thus, B is better than C. Finally, C is better than A with

respect to R1 and incommensurate with A in the other two respects. Thus, C is better than A. In sum, A is better than B, B is better than C, and C is better than A.

It is easy to see how this argument can be implemented to argue against Strong Pseudo-Dominance. We only need to find a supervenience set that contains three respects and let R1, R2, and R3 be these three respects. Strong Pseudo-Dominance then implies that X is overall better than Y if X pseudo-dominates Y in {R1, R2, R3}, so the cyclicity argument applies. On many views, it is not hard to find a supervenience set with three respects. On the total view, we could use {total human wellbeing, total shrimp wellbeing, total wellbeing of all other individuals}. On egalitarianism, we could use {total human wellbeing, total non-human wellbeing, equality}. On a non-welfarist axiology, we could use {total wellbeing, knowledge, beauty}. And if there are only supervenience sets with more than three respects, we can specify R1, R2, and R3 such that they form a *subset* of a supervenience set while stipulating that the compared things don't differ in any other respect of that supervenience set.⁵³ Given this stipulation, if X pseudo-dominates Y in {R1, R2, and R3}, X also pseudo-dominates Y in a supervenience set (and is therefore overall better than Y according to Strong Pseudo-Dominance), so the cyclicity argument still applies. For example, if any of the above sets are only subsets of more encompassing supervenience sets that additionally include the respects of suffering and achievement, we could stipulate that A, B, and C don't differ in the latter respects. Similarly for the versions of Strong Pseudo-Dominance that specify a supervenience set. For example, to argue against individual-wise pseudo-dominance, we can specify R1, R2, and R3 such that they

⁵³ Arguably, there is a supervenience set with no more than three respects on *any* axiology: we could always reduce the number of respects in a supervenience set by replacing several respects by their conjunction, which is itself a (conjunctive) respect in which things can be better or worse.

are the wellbeing levels of the only individuals whose wellbeing differs between A, B, and C (see Hare 2013, 176-177).⁵⁴

In principle, the argument could be extended in a similar way to Weak Pseudo-Dominance: we could stipulate that A, B, and C don't differ in any respect *at all* besides in R1, R2, and R3 – not just in no other respect in a given supervenience set, but in no other respect in *any* supervenience set. The problem is that this stipulation is in conflict with the other stipulation that A, B, and C are incommensurate in some respects. Both stipulations could only hold simultaneously if things could be incommensurate in a respect without each thing being better than the other in some other respect. But as we have seen in the previous subsection, it is questionable whether any such cases exist.⁵⁵ None are demonstrated by the authors who employ the cyclicity-argument since they are (or seem to be) only concerned with versions of Strong Pseudo-Dominance. Normally, it is just *stipulated* that there is incommensurateness in some of the three respects or the incommensurateness is made plausible by cases in which each thing is better than the other in some respect, similarly to the cases discussed in the previous section (e.g. Hedden & Muñoz 2024, 300).

3.3 Crosswise Sweetening

The two previous objections are based on cases in which pseudo-dominance principles seem to imply betterness even though there is no betterness. According to another

⁵⁴ Nebel (2019, 139) and Hedden & Muñoz (2024, 300) use the same case to argue against dimension-wise pseudo-dominance. Again, I am sceptical of this argument if value dimensions are supposed to be something more specific than respects in which things can be better or worse.

⁵⁵ The focus in the last section was on *overall* incommensurateness, not incommensurateness *in a respect*. But one can be easily translated into the other by making the value bearers, X and Y, more or less complex. For example, on the total view, if two welfare distributions are incommensurate with respect to the total wellbeing of a subpopulation, then two distributions that only concern this subpopulation will be incommensurate overall. And if there are two distributions that are incommensurate overall, there will be two distributions concerning larger populations that are incommensurate with respect to the total wellbeing of that subpopulation.

objection, closely related principles imply in some cases that there is *no* betterness even though there *is* betterness. These are *negative* dominance principles, which state that X is *not* better than Y if, at least within some set of respects, X is not better than Y in any respect.⁵⁶ Put differently, and to emphasize their similarity with pseudo-dominance principles, negative dominance principles state that X is *not worse* than Y if, at least within some set of respects, X is not worse than Y in any respect. Analogously to pseudo-dominance principles, we can distinguish between a weak and a strong form of negative dominance.

Weak Negative Dominance: For all X and Y, if X is not worse than Y in any respect, then X is not worse than Y overall.

Strong Negative Dominance: For all X and Y, if there is a supervenience set that does not contain any respect in which X is worse than Y, then X is not worse than Y overall.

Plausibly, pseudo-dominance principles could only be true if the corresponding negative dominance principles are true. In particular, Weak Pseudo-Dominance is plausibly only true if Weak Negative Dominance is true. If Weak Negative Dominance was false, X could be worse than Y even though X is not worse than Y in any respect. But if being not worse in any respect is not sufficient for being not worse overall, it seems also not sufficient for being *better* overall that something is not worse in any respect while being better in *some* respect. After all, this respect could be any minor respect and the betterness in this respect could be ever so slight. Of course, even if Weak Negative Dominance was false, it could be that any X that is not worse than Y in any respect is at most *slightly* worse than Y. In

⁵⁶ The name “negative dominance” is due to Lederman (2023) who uses it to denote the principle that a lottery is not better than another lottery if none of its possible outcomes is better than any of the possible outcomes of the other lottery. In my terminology, this principle is (plausibly) an instance of Strong Negative Dominance.

that case, the additional respect in which X is better than Y, which the condition of Weak Pseudo-Dominance requires, would not have to be a respect in which X is much better than Y – being only slightly better in that respect could suffice for overall betterness. But it seems plausible that we can always find a respect in which X could be better than Y in such a minor way that it would not suffice to make X better overall. For example, X could be better than Y in the respect that X contains one additional minute (or second, or millisecond, ...) of additional pleasure. However slightly some X may be worse than Y, not *any* additional duration of pleasure will make it overall better than Y.

It seems, therefore, that any objection against Weak Negative Dominance will equally affect the plausibility of Weak Pseudo-Dominance. Now, Steele (2022) argues against a negative dominance principle that superficially resembles Weak Negative Dominance. According to this principle, “an option is *not* worse than another option, and thus (criterion-wise) competitive, if, under every *criterion*, its outcome is *not* worse than that of the other option” (Steele 2022, 235). Steele argues that this principle, which she calls “(Criterion-Wise) Competitive”, implies that a sweetened option is not better than an unsweetened option if the sweetening is applied in a certain ‘crosswise’ manner.⁵⁷ To demonstrate this, Steele uses the following example, where Box 1 and Box 2 are transparent boxes between which we can choose that contain gifts for Pablo and Penelope. Gifts A and B are incommensurate for both Pablo and Penelope, who gain exactly the same welfare as one another from each gift, and A+ and B+ are sweetened versions of A and B that additionally include a dollar. Since A+ is only slightly better than A, A+ is also incommensurate with B for both Pablo and Penelope, and since B+ is only slightly better than B, B+ is also incommensurate with A for both of them.⁵⁸

⁵⁷ I take the designation “crosswise sweetening” from Wlodek Rabinowicz.

⁵⁸ This is a version of Hare’s (2010) “opaque sweetening” case, where we are uncertain about the contents of the two boxes.

	Box 1	Box 2
Pablo	A	B+
Penelope	B	A+

Box 2 is a crosswise sweetened version of Box 1: Box 2 contains slightly improved versions of the gifts in Box 1, but the receivers of the gifts are swapped. Steele argues that, intuitively, choosing Box 1 is worse than choosing Box 2, but that (Criterion-Wise) Competitive implies that Box 1 is not worse than Box 2.⁵⁹ That is because Box 1 is not worse than Box 2 for both Pablo and Penelope: A and B+ are incommensurate for Pablo, and B and A+ are incommensurate for Penelope. Steele concludes that, since the two options don't differ in any other relevant criteria, Box 1 is not worse than Box 2 under any criterion and hence, according to (Criterion-Wise) Competitive, not worse overall.

Although (Criterion-Wise) Competitive superficially resembles Weak Negative Dominance, Steele's argument is best understood as an argument against (a version of) Strong Negative Dominance.⁶⁰ Given Steele's assumptions, it is plausible that Box 1 is not worse than Box 2 in any respect within a supervenience set. But Box 1 is clearly worse than Box 2 in some respects. The fact that gifts A and B are incommensurate for both individuals is plausibly explained by the fact that each gift is better than the other in some respect. In fact, to make their incommensurateness plausible, Steele asks us to imagine that "A is a

⁵⁹ Steele's claim that Box 2 is better than Box 1 can be supported by the following argument: Box 2 is as good as a third box that contains the same gifts, A+ and B+, but assigns A+ to Pablo and B+ to Penelope. This third box is clearly better than Box 1: it is better for both Pablo and Penelope. But then, Box 2, which is as good as the third box, also needs to be better than Box 1. This argument is analogous to Hare's (2010, 240) argument from "Prospects Determine Permissibility" and Bader's (2018) argument from Stochastic Dominance in the opaque sweetening case.

⁶⁰ We could understand the "criteria" in (Criterion-Wise) Competitive as particular respects, analogously to value dimensions, that jointly form a supervenience set.

ticket to the opera while B is a ticket to an A-league sports game” (2022, 240). Obviously, each of these is better than the other in some respect. Moreover, the very fact that the gifts in Box 2 are sweetened is a respect in which Box 2 is better than Box 1. In the concrete imagined case, Box 2 is better than Box 1 with respect to the monetary gain of both Pablo and Penelope. Thus, Weak Negative Dominance does not imply that Box 2 is not better than Box 1.⁶¹

4. Incommensurateness Requires Value Conflict

I have argued that various objections against pseudo-dominance principles don’t affect Weak Pseudo-Dominance because these objections are based on cases of incommensurateness in which each of the compared things is better than the other in some respect. In such cases, Weak Pseudo-Dominance simply does not apply. In this section, I will argue that, at least within a suitably restricted domain of evaluands, it is plausible that any instance of incommensurateness is grounded in some value conflict. I will also argue that if this is correct, Weak Pseudo-Dominance is true.

4.1 Weak and Strong Incommensurateness Requirements

I will distinguish between two requirements for incommensurateness that differ in whether they require that incommensurateness is *fully* grounded in value conflicts.

Weak Incommensurateness Requirement: For all X and Y that are commensurate in at least one respect, if X and Y are incommensurate in some respect R (or

⁶¹ In fact, Box 2 *dominates* Box 1 in a supervenience set containing the respects in which gifts A and B differ and the (monetary) respect in which the sweetened gifts in Box 2 are better (see Francis & Gustafsson 2025, 8-9).

incommensurate overall), then there is a non-empty set S1 that only contains respects in which X is better than Y, there is a non-empty set S2 that only contains respects in which Y is better than X, and the facts about how much X is better than Y in the respects in S1 and about how much Y is better than X in the respects in S2 at least partially ground the fact that X and Y are incommensurate in R (or incommensurate overall).

Strong Incommensurateness Requirement: For all X and Y that are commensurate in at least one respect, if X and Y are incommensurate in some respect R (or incommensurate overall), then there is a non-empty set S1 that only contains respects in which X is better than Y, there is a non-empty set S2 that only contains respects in which Y is better than X, and the facts about how much X is better than Y in the respects in S1 and about how much Y is better than X in the respects in S2 fully ground the fact that X and Y are incommensurate in R (or incommensurate overall).

Both principles entail that incommensurateness between X and Y requires that there is at least one respect that favours X and at least one respect that favours Y. They moreover require that there is a *grounding* relationship between some sets of such counteracting respects and incommensurateness:⁶² that some of these respects favour X over Y to certain extents while the other respects favour Y over X to certain extents grounds the fact that X

⁶² I often simply say that certain (conflicting) respects ground incommensurateness rather than saying that certain *facts* about these respects ground the *fact* that there is incommensurateness. This is mainly for simplicity and greater intuitiveness. As the subsequent sentence in the main text illustrates, these claims can be translated into more complicated claims about which facts ground facts about incommensurateness. Maybe the grounding relation holds, strictly speaking, only between facts (see Rosen 2010, Audi 2012).

and Y are incommensurate (in some respect or overall).⁶³ This additional requirement makes the principles logically stronger but barely less intuitively plausible: that there is such a grounding relationship is what motivates the thought that incommensurateness requires value conflicts in the first place.⁶⁴ For example, the options of saving the wedding album and of saving the first edition *Meditations* are incommensurate *in virtue* of the fact that the wedding album is much better with respect to sentimental value while the first edition *Meditations* is much better with respect to cultural value. And if two careers are incommensurate for some person (and two states of affairs are incommensurate with respect to this person's career), that is *because* both careers have some significant advantages that cannot be precisely weighed against each other.

As these examples make plausible, it seems that when value conflicts ground incommensurateness, value conflicts not only form a *partial* ground of incommensurateness but *fully* ground incommensurateness. For example, that the options of saving the wedding album and saving the *Meditations* are incommensurate seems to be fully grounded in the value conflict between cultural value and sentimental value. Of course, not each instance of incommensurateness is fully grounded in a conflict between only two respects: there may be many counteracting respects, and the incommensurateness may be fully grounded only in the conflict between all of these counteracting respects –

⁶³ The addition of “to certain extents” and of “how much” in the formulations of the two incommensurateness requirements is necessary because not every conflict between given respects grounds incommensurateness. Even if conflicts between respects R1 and R2 *sometimes* ground incommensurateness, they may not ground incommensurateness if X is *much* better than Y with respect to R1 and Y is only *slightly* better than X with respect to R2. At least for the *Strong* Incommensurateness Requirement, we thus need *cardinal* axiological facts about how the conflicting respects relate to X and Y. One could argue that even such cardinal facts are not enough for *full* grounding and that we additionally need the fact that the respects that favour X and Y cannot be *precisely weighed* against each other. Such facts are plausibly part of a full *explanation* for incommensurateness; I find it less plausible that they are also part of what *grounds* the incommensurateness. As far as my argument is concerned, nothing seems to hinge on how this metaphysical issue is decided and so I prefer the simpler principle that does not include this fact.

⁶⁴ Strictly speaking, the additional requirement does make the principles less plausible because, as I will argue in the next section, it seems that *whenever* X is better than Y in some respect, Y is better than X in some respect. If this is right, the additional requirement saves the principles from being almost *vacuously* true – and thereby, of course, makes them less plausible.

this is why the principles refer to *sets* of respects rather than individual respects. But once we have listed, for example, all the respects that favour the career as a musician and all the respects that favour the career as a painter, and we have moreover specified the extent to which each respect favours one of the two careers, it seems that we are not missing anything that grounds the incommensurateness of the two careers. In particular, we don't additionally need any respects in which the compared things are *incommensurate*. As we have seen, there plausibly are such respects, such as the respect of sentimental-cum-cultural value, but these respects are themselves grounded in value conflicts between counteracting respects.⁶⁵

It seems plausible, then, that *when* value conflicts ground incommensurateness, they *fully* ground incommensurateness. Accordingly, it seems plausible that *if* the Weak Incommensurateness Requirement holds, the Strong Incommensurateness Requirement holds too. For now, however, the difference between the two principles is not important.⁶⁶ The weaker principle is all we need for Weak Pseudo-Dominance. I will show this in the next subsection and argue for the Weak Incommensurateness Requirement in the remaining subsections.

4.2 An Argument for Weak Pseudo-Dominance

There is a straightforward argument from the Weak Incommensurateness Requirement to Weak Pseudo-Dominance. We only need to add a strict dominance principle that is analogous to Weak Pseudo-Dominance.

⁶⁵ How about respects in which X and Y are equally good? Equal goodness in a respect can make a counterfactual difference as to whether X and Y are incommensurate: it could be that X and Y are incommensurate overall and equally good in a respect, but that X and Y would not be incommensurate overall if they were not equally good in that respect. But intuitively this does not suffice to make the respect in which X and Y are equally good a *ground* of the incommensurateness: the respect just does not undermine the incommensurateness.

⁶⁶ The difference will only become relevant later in the paper when I defend a stronger version of Weak Pseudo-Dominance, which I will call "Generalized Pseudo-Dominance".

Weak Dominance: For all X and Y, if X is better than Y in some respect and X is at least as good as Y in every respect, then X is better than Y overall.

Analogously to the distinction between Weak Pseudo-Dominance and Strong Pseudo-Dominance, Weak Dominance can be contrasted with a stronger principle that relativizes the dominance relation to supervenience sets.

Strong Dominance: For all X and Y, if there is a supervenience set that contains a respect in which X is better than Y and that only contains respects in which X is at least as good as Y, then X is better than Y overall.

Unlike Strong *Pseudo*-Dominance, Strong Dominance does not create any problems with incommensurateness, so it seems like a plausible principle. I will later argue that there are still reasons to doubt Strong Dominance and that we need to add at least one qualification to the principle. But none of this concerns Weak Dominance, which is about as weak as dominance principles can get. It seems obviously true.

The following argument shows that the conjunction of the Weak Incommensurateness Requirement and Weak Dominance implies Weak Pseudo-Dominance. (For readability, I have skipped the quantifier “for all X and Y” throughout the argument.)

- (1) The Weak Incommensurateness Requirement is true.
- (2) If the Weak Incommensurateness Requirement is true, X and Y are not incommensurate in any respect if X is better than Y in some respect but X is not worse than Y in any respect.
- (3) If X is not worse than Y in any respect, and X and Y are not incommensurate in any respect, then X is at least as good as Y in every respect.

- (4) Thus, if X is better than Y in some respect and X is not worse than Y in any respect, then X is at least as good as Y in every respect. (From (1), (2) and (3).)
- (5) Weak Dominance: If X is better than Y in some respect and X is at least as good as Y in every respect, then X is better than Y overall.
- (6) Weak Pseudo-Dominance: If X is better than Y in some respect and X is not worse than Y in any respect, then X is better than Y overall. (From (4) and (5).)

The only substantial premises of the argument are the Weak Incommensurateness Requirement (1) and Weak Dominance (5) – premises (2) and (3) are analytically true and (4) follows from premises (1) to (3). The argument thus shows that Weak Pseudo-Dominance is equivalent with Weak Dominance if the Weak Incommensurateness Requirement holds. As I commented above, Weak Dominance seems obviously true, so all hinges on the Weak Incommensurateness Requirement.

4.3 The Basic Case for the Weak Incommensurateness Requirement

Weak Pseudo-Dominance is true if the Weak Incommensurateness Requirement is true. But why should we believe in the latter? On the most general level, the requirement provides the basis for a unified account of incommensurateness – one according to which all cases of incommensurateness are grounded in some value conflict – while leaving a lot of room for incommensurateness that includes its most plausible instances. In particular, it is compatible with all the paradigmatic cases of incommensurateness from the previous section that are often referenced to make the existence of incommensurateness plausible in the first place. The same could be said about other stock examples from the incommensurateness literature. For example, what makes the choice for Sartre's (1975) student between joining the French forces in England or caring for his sick mother at home

so hard is that each option is better in some significant way. If the student knew that he would be completely useless as a member of the French forces, or that he could support them equally well from home while caring for his mother, the choice wouldn't be a hard one: it would be better for him to stay home.

We can distinguish between two types of value conflicts: *inter*-value conflicts and *intra*-value conflicts. Inter-value conflicts are conflicts between different fundamental values, such as pleasure and achievement, or welfare and justice. Two things exhibit an inter-value conflict if each of them is better than the other with respect to some fundamental value. Such value conflicts are the most commonly acknowledged source of incommensurateness. They seem to explain the incommensurateness in many of the cases that I have covered so far. For example, the respects that favour each option of Sartre's student seem to correspond to fundamentally different values, such as freedom and honour on the one hand and love and compassion on the other hand (see Broome 2004, 168).

Besides such inter-value conflicts, there are also *intra*-value conflicts, which concern different aspects of the same value. Two things exhibit an intra-value conflict if each is better than the other in some respect and both these respects (and the properties that determine them) derive their value from the same fundamental value. It is less commonly acknowledged that intra-value conflicts can lead to incommensurateness, but it seems plausible to me that they can (conditional on the existence of incommensurateness, at least). Consider the options of organising two different large-scale events: a football match or a concert. These two events would attract very different people and cause them to have pleasurable experiences from different sources, with different intensities and durations. Consequently, one of them is better with respect to Amy's pleasure while the other is better with respect to Tomi's pleasure, one is better with respect to pleasure from music while the other is better with respect to pleasure from sports, and so on. But which event is better with respect to pleasure overall? It seems possible that neither is better than the other with

respect to pleasure and that they are also not equally good with respect to pleasure. One could explain this intuition on the basis of Mill's (1863) view that pleasure has different *qualitative* aspects (cf. Chang 1997, 16–17). But the explanation does not have to be that theoretically intricate. It could just be the case that neither event produces *more* pleasure than the other. The conflict between different merely *quantitative* aspects of pleasure could, so to speak, result in *descriptive* incommensurateness with respect to pleasure, and this descriptive incommensurateness could ground the evaluative incommensurateness: neither event is at least as good as the other with respect to pleasure because neither event produces at least as much pleasure as the other event.

The Weak Incommensurateness Requirement allows that both inter-value conflicts and intra-value conflicts can ground incommensurateness. This reaffirms my point that the Weak Incommensurateness Requirement leaves a lot of room for incommensurateness. Even value monists, such as hedonists, could believe in incommensurateness. And since intra-value conflicts, at least, are extremely common, incommensurateness could be a ubiquitous phenomenon as far as the Weak Incommensurateness Requirement is concerned.

4.4 Incommensurateness due to Ontological Differences

It seems that there are some trivial cases of incommensurateness that are not grounded in any value conflict. It seems, for example, that Aristotle is incommensurate with climate change, and that the colour blue is incommensurate with your table.⁶⁷ But, at least on the face of it, there seems to be no value conflict between Aristotle and climate change, or between the colour blue and your table, that explains their incommensurateness. It is not

⁶⁷ If this sounds odd, here is a reminder that I use “incommensurate” in the entirely negative sense of being neither better nor worse nor equally as good.

clear what exactly we should say about such cases. Their incommensurateness seems to be somehow explained by the fact that the evaluands belong to different ontological categories. But why does membership in different ontological categories (sometimes) create incommensurateness?

One possible answer is that only things from one ontological category are *value bearers*, so that any comparisons between things from different ontological categories involve at least one thing that is not a value bearer. Specifically, it has some plausibility that only *states of affairs* are value bearers. The apparent incommensurateness in the above cases could then be explained by the fact that they involve things that are not states of affairs. Another possible explanation is that if the evaluands belong to different categories, they sometimes have different *standards of evaluation*. For example, it could be that both people and events are value bearers, but that what makes people better or worse are their virtues and vices while what makes events better or worse are their effects on welfare subjects. Consequently, even very good people like Aristotle won't be better than very bad events like climate change: the virtues of Aristotle don't matter in a comparison with climate change and the bad effects of climate change don't matter in a comparison with Aristotle.⁶⁸ Finally, one could evoke a notion of *comparability* that is broader than that of commensurateness, and argue that things from different ontological categories are (often) not even comparable.⁶⁹ Given that comparability is a broader notion than

⁶⁸ There is an interesting variation of this explanation: The same things can be evaluated based on different standards of evaluation. Normally, when we compare two members from some (not-very-broad) category, such as two climate change scenarios, it is clear what the relevant standard of evaluation is. But when we compare two rather different things, such as a person and climate change, and there is no additional context, it seems *indeterminate* what the relevant standard of evaluation is. Consequently, it is also indeterminate which of the evaluands is better. Unless the relevant standard of evaluation is stipulated, there is just no fact of the matter how Aristotle and climate change compare. On this account, there is no incommensurateness in the above cases after all but indeterminacy.

⁶⁹ See Chang (2002) who argues that there is an intuitive, non-stipulative sense of comparability that is broader than that of commensurateness. In that sense, comparable things could exhibit a non-standard value relation such as parity (Chang 2002) or rough equal goodness (Griffin 1986, 80; Parfit 1984, 431).

commensurateness, it would follow trivially that things from these different categories are also incommensurate.

These explanatory approaches suggest different scope limitations of the principle that incommensurateness requires value conflict: it holds only for *value bearers*, or things with the *same evaluative standards*, or *comparable* things. The Weak Incommensurateness Requirement entails a scope limitation that is largely neutral between these different suggestions and that can even be understood as a way of making some of them more precise: it is limited to such X and Y that are *commensurate in at least one respect*. Plausibly, being commensurate with something else in some respect – i.e. being better or worse or equally as good in this respect – is sufficient for being a value bearer. In fact, this condition could be taken to make the notion of a value bearer more precise: a value bearer is something that is better or worse or equally as good as something else in at least some respect. Commensurateness in some respect also ensures that the evaluative standards of the compared things at least overlap. If X and Y are commensurate in some respect, it cannot be the case that X and Y have no evaluative criterion in common. Finally, commensurateness in some respect seems sufficient for comparability. If X is at least as good as Y in some respect (or vice versa), it seems intuitively wrong to say that X and Y are incomparable. Again, the condition even provides a plausible precisification of what (in-)comparability is: two things are incomparable if and only if there is no respect in which one of them is at least as good as the other.

Consider again the above examples. Aristotle and climate change seem to be not only incommensurate overall but also incommensurate in all respects. They seem to be (in that sense) just *incomparable*. Thus, the Weak Incommensurateness Requirement is compatible with their incommensurateness. In the other example, it seems plausible that one of the evaluands is not even commensurate in any respect with *anything* else (and is thus no *value bearer*): there seems to be nothing relative to which the colour blue is better

or worse or equally as good. *A fortiori*, it seems that the colour blue is not commensurate with a *table* in any respect.⁷⁰

I conclude that the scope limitation of the Weak Incommensurateness Requirement is a plausible solution to the problem of differing ontological categories, as exemplified by the above cases. But not much hinges on whether this particular scope limitation is sufficient. There are many related possible scope limitations – such as those suggested by the explanatory approaches mentioned above – that could be applied instead. Some of these would require that also the scope of Weak Pseudo-Dominance is (somewhat) restricted. But many such restrictions would not make Weak Pseudo-Dominance significantly less interesting. For example, one particularly safe way to avoid the problem of differing ontological categories would be to limit the Weak Incommensurateness Requirement to one category of things at a time, such that we have one requirement for states of affairs, one for people, one for actions, and so on. Accordingly, Weak Pseudo-Dominance would need to be equally split up in many principles, each referring to X and Y of only one such category. Unless we were hoping to use pseudo-dominance principles to compare things across such broad categories of things, the set of all these specified principles seems barely less interesting than the general principle.

4.5 Incommensurateness due to Non-Existence of Determinates

Even when we compare members of the same ontological category, such as two states of affairs, it may seem that there are trivial instances of incommensurateness that are not

⁷⁰ At least, this seems plausible on the assumption that the evaluands in these cases are indeed incommensurate overall. If you think that even a table and a colour could be commensurate overall, perhaps because this follows from the general logic of comparatives (see Dorr, Nebel & Zuehl 2023), or that it could be *indeterminate* whether they are commensurate overall (see the footnote above), then it also seems (not determinately false) that they are commensurate in some respects. This is of course compatible with the Weak Incommensurateness Requirement.

grounded in value conflicts. Cusbert & Kath (2019, 1696) argue that things can be incommensurate in a respect simply because they don't both *register* on that respect. To illustrate the notion of (not) registering on a respect, they give the following examples: “the empty outcome does not register with respect to average wellbeing; and those outcomes in which Anna does not exist arguably do not register with respect to Anna’s wellbeing” (ibid.). Here is a way of making the notion of registering on a respect more precise: respects are *determinable* properties that are determined by more specific, *determinate* properties, where each thing that instantiates a determinable property also instantiates exactly one of the properties that determine it (cf. Nebel 2015, 11).⁷¹ For example, average wellbeing can be determined by different specific amounts of average wellbeing (which we could denote by numbers) and everything that has an average wellbeing has exactly one such specific amount. When something does not have any determinate property that determines a given respect (and hence does not have the determinable property that *is* the respect), we can say that it does not “register” on that respect. For example, the empty outcome does not register on the respect of average wellbeing in the sense that it does not have any particular average wellbeing that determines that respect.

Cusbert’s & Kath’s (2019) view is that whenever something does not register on a respect, it is incommensurate with anything else in this respect. That is, whenever something does not have any property that determines a given respect, that thing is incommensurate with anything else in this respect. They seem to take this to be obviously true. According to them, saying that X is better than Y with respect to R when X or Y don’t register on R is “akin to a category error” (2019, 1697). I am unsure whether this is correct. Even when X or Y don’t have any R-determinate, we can say the following about how they

⁷¹ Properties that determine the same determinable property are thus *mutually incompatible*. This distinguishes the determinable-determinate relation from other specification relations, such as that between disjunctions and their disjuncts (see Wilson 2023).

compare with respect to R: one of them has a certain R-determinate while the other does not have any R-determinate. For example, the empty outcome does not have any average wellbeing while an outcome in which everyone is suffering has a negative average wellbeing. Clearly, the empty outcome is better than the other outcome in *some* respect in virtue of this difference. The question is whether the respect in which the empty outcome is better in virtue of this difference is the respect of *average wellbeing*. One could argue that, strictly speaking, the respect in which the empty outcome is better than the other outcome is that of *whether something has a negative average wellbeing*, which can be determined by the properties in the set {has a negative average wellbeing; does not have a negative average wellbeing}. On the other hand, it does seem natural to me to say that the empty outcome is better than the other outcome “with respect to average wellbeing”: it is better in this respect because it has no average wellbeing while the outcome has a negative average wellbeing.

Ultimately, nothing important seems to depend on how this issue is decided. If there is incommensurateness between two things whenever one of them does not register on a respect, such incommensurateness would not only be ubiquitous but also evaluatively irrelevant. Any two things would be incommensurate in countless respects that we don’t need to consider to determine their overall comparative value. Consider two populations that contain only one individual, Amy, who is slightly better off in one population than in the other. Obviously, the population in which she is better off is better overall. But the two populations are incommensurate with respect to Bob’s wellbeing, Tomi’s wellbeing, Mia’s wellbeing, and so on. Normally, when two things are incommensurate in one respect (let alone in *countless* respects) and there is only one respect in which one of them is slightly better than the other, they are overall incommensurate. But here, the incommensurateness

is not ‘insensitive to sweetening’ – it can be ignored similarly to how we normally ignore respects in which things are equally good.⁷²

Arguably, this evaluative characteristic shows that the assumed incommensurateness really is equal goodness. But be this as it may, if there is incommensurateness in virtue of not registering, this does not affect the plausibility of Weak Pseudo-Dominance since it *behaves* just like equal goodness. I will therefore ignore such incommensurateness and (for simplicity) assume that it does not exist.

4.6 Controversial Cases of Incommensurateness

There remains the distinct question whether it is sometimes the case that two things are incommensurate in a respect in virtue of the fact that one of them determines this respect in a *certain* way while the other does not have any property that determines the respect. Above, I have given an example in which this is not the case: the empty outcome is better than the outcome with suffering people in some respect (whatever *exactly* this respect is) in virtue of the fact that the empty outcome does not have any average wellbeing while the other outcome has a negative average wellbeing. But one might hold that if the average wellbeing in the non-empty outcome was instead slightly positive, the two outcomes would be incommensurate with respect to (something like) average wellbeing. More commonly, the view has been defended that two outcomes can be incommensurate in virtue of the fact that an individual who has a good life in one outcome does not exist in the other outcome (e.g. Rabinowicz 2009; Frick 2017). This view is not obviously in conflict with the Weak Incommensurateness Requirement. *Prima facie*, it could be that whenever only one of two outcomes contains some happy individual, each outcome is better than the other in some

⁷² In fact, each of the two populations would be incommensurate *with itself* in all of these respects. And still, each population is overall as good as itself – despite being incommensurate with itself in so many respects.

respect. But I will later argue that, in fact, there are some possible cases in which the outcome with an additional happy person is not worse in any respect. If this is right, there will be a real conflict with the Weak Incommensurateness Requirement.

What this shows is that the Weak Incommensurateness Requirement has non-trivial implications. But it doesn't show that the Weak Incommensurateness Requirement is overall implausible. We have now moved away from paradigmatic cases of incommensurateness that are hard to deny to controversial cases of incommensurateness that have some plausibility. It is to be expected that a principle like the Weak Incommensurateness Requirement that unifies cases of incommensurateness will be incompatible with some cases of the latter kind.

At this point, it is important to consider the whole dialectical situation, which involves not only putative cases of incommensurateness and the Weak Incommensurateness Requirement, but also Weak Pseudo-Dominance. Weak Pseudo-Dominance is in itself an intuitively compelling principle. The Weak Incommensurateness Requirement was not supposed to convey plausibility to it in first place. Rather, it was supposed to disperse the worry that Weak Pseudo-Dominance requires an implausibly restrictive view on incommensurateness. This aim is not undermined by the fact that the Weak Incommensurateness Requirement rules out some plausible-but-controversial cases of incommensurateness. To achieve that aim, the Weak Incommensurateness Requirement does not have to be an uncontroversial principle; it is enough that it is one of several plausible views. The view that incommensurateness can also be grounded in facts of the kind that some individuals exist only in one of two compared outcomes could be another plausible view. There will still be a plausible view on incommensurateness that is compatible with Weak Pseudo-Dominance, so the objection against it is dispersed.

I conclude that objections from incommensurateness don't significantly reduce the plausibility of Weak Pseudo-Dominance. Unfortunately, there is a different problem for Weak Pseudo-Dominance that is more severe.

5. The Application Problem

Normally, when we compare the value of two things, each of them is better in at least some minor respect. Even in easy choices in which one option is clearly worse, there is normally *something* to be said in favour of that option. This means that Weak Pseudo-Dominance will be rarely applicable in practice. But that by itself would not be a serious problem. Weak Pseudo-Dominance shares this characteristic with many other principles, such as *ex ante* Pareto and *state-wise dominance*.⁷³ These principles rarely apply in real cases, and when they apply, their implications may seem obvious anyway. But that doesn't mean that they cannot do interesting work. For one, they can be applied to hypothetical cases to assess moral theories, which have implications for real and hypothetical cases alike. Moreover, their conjunction with other principles often does have interesting implications even for real cases. Accordingly, even if they cannot do much interesting work by themselves, they can do so together with other principles.

However, there is a distinct application problem for Weak Pseudo-Dominance. Unlike in the case of *ex ante* Pareto or *state-wise dominance*, it is difficult to even find *hypothetical* cases in which Weak Pseudo-Dominance is applicable. There is a systematic reason that seems to make its application impossible. Here is a first sketch of the problem: Weak Pseudo-Dominance can only be applied if one of the compared things, X, is better

⁷³ Roughly, *ex ante* Pareto states that an option is better overall if, in expectation, it is at least as good for each individual and better for some individual, and *state-wise dominance* states that an option is better overall if it is at least as good in each possible state of nature and better in some possible state of nature.

than the other, Y, in some respect. This requires, of course, that X and Y differ in that respect: X and Y must not have the same properties that determine the respect – otherwise they would be equally good in the respect. The problem is that, at least normally, this descriptive difference can be used by *another* respect, which is specified by one of the two different properties of X and Y, to single out some (possibly unrelated) bad feature of X or good feature of Y. Since this other respect is thus tailored to a bad feature of X or a good feature of Y, it will favour Y over X. Consequently, each of X and Y will be better than the other in some respect if one of them is better than the other in some respect.

Here is an example: let P and P+ be two populations containing the same one billion people that only differ in that these people are better off in P+ than in P. In P+ they have the high wellbeing level 10 while in P they have the somewhat lower but still high wellbeing level 9. This is a respect in which P+ is better than P, and it seems that P+ and P could be stipulated such that P+ is not worse than P in any respect. If Weak Pseudo-Dominance is ever applicable, it should be applicable in this case. However, because only P contains people with wellbeing 9, there are some respects that reflect only the value of the wellbeing in P but not the value of the wellbeing in P+. Consider, for instance, the following respect: total wellbeing of everyone with a wellbeing of (at most) 9. In P+, the total wellbeing of this subpopulation is 0 since no one has a wellbeing of 9 or lower. In P, the total wellbeing of this subpopulation is 9 billion since there are one billion people with wellbeing 9. Given that total wellbeing matters, this is a respect in which P+ is worse than P.⁷⁴ The same holds for other respects that make reference to people with wellbeing 9. For example, if these people make any achievements in P, P+ will also be worse than P with respect to the

⁷⁴ If this seems doubtful, note that total wellbeing supervenes on sets that contain this respect along with similar respects, such as {total wellbeing of people with a wellbeing below 9, total wellbeing of people with wellbeing 9, total wellbeing of people with a wellbeing higher than 9}, and note that total wellbeing strictly increases with each respect in that set, just as total wellbeing strictly increases with the respects in the set {total human wellbeing, total non-human wellbeing}.

achievements of people who have a wellbeing of (at most) 9.⁷⁵ Of course, there are corresponding respects that only reflect the welfare value of P^+ (such as total wellbeing of people with wellbeing 10) and in which, accordingly, P^+ is better than P . But the applicability of Weak Pseudo-Dominance requires that there is no respect in which P^+ is worse than P . That is not the case. Thus, Weak Pseudo-Dominance is not applicable.

It seems that the same could be said about virtually all other respects in which something can be better than another thing: the descriptive difference that grounds the betterness in a given respect R can be employed by another respect to single out some valuable feature of the thing that is worse with respect to R (or some bad feature of the thing that is better with respect to R).⁷⁶ If the other respect is sufficiently tailored to this valuable feature, such that the respect doesn't additionally reflect some other evaluatively relevant features of X or Y , the thing that is worse with respect to R will be better in the other respect. And it seems that such tailoring is virtually always possible. In the above example, not much tailoring was needed since P^+ does not contain any people with wellbeing 9. So it was sufficient to compare P and P^+ with respect to the total wellbeing of people with wellbeing 9 to single out the welfare value of P . But the same could be done with a more tailored respect if we assume that P^+ contains one billion people with wellbeing 9 in addition to the one billion people with wellbeing 10. Given this new assumption, P is not better than P^+ with respect to total wellbeing of people with wellbeing 9, but P is better than P^+ in the more tailored respect of total wellbeing of people with wellbeing 9 who are at least as well off as everyone else.

⁷⁵ The same holds for some respects that only refer to people with wellbeing 10 (or higher). For example, if the lives of the people in P^+ contain any suffering, P^+ will be worse than P with respect to the total suffering of people with wellbeing 10 (or higher).

⁷⁶ To give another example, the fact that one action benefits a given person to a higher degree than another action means that, if all else is equal, the latter action confers more benefits up to that degree.

There remain some open questions about the scope of the application problem. Perhaps, there are some respects after all in which something can be better without being worse in another respect.⁷⁷ But even if that is the case, the applicability of Weak Pseudo-Dominance is still severely limited. The above example is a case in point. If Weak Pseudo-Dominance cannot even be applied in such an idealized case, the principle barely seems worthwhile.

6. Solving the Application Problem

The application problem shows that Weak Pseudo-Dominance is not (or barely) applicable as it stands. To make it applicable, we need to allow that the pseudo-dominant thing is worse in some respects. That seems to contradict the very idea of Weak Pseudo-Dominance: that something is better if there is something about it that is better and *nothing* about it that is worse. But perhaps the core of that idea can be maintained if we only allow that the pseudo-dominant thing is worse in certain respects that are intuitively negligible or for which it is *trivially* the case that the pseudo-dominant thing is worse with respect to them. Perhaps, these respects are simply not worth mentioning under normal circumstances, but if pointed out, anyone who is reasoning along the lines of Weak Pseudo-Dominance would be happy to acknowledge them. This seems plausible for the respects with which I illustrated the application problem. Intuitively, for instance, the respect of total wellbeing of people with wellbeing 9 is negligible in my example and cannot undermine any relevant pseudo-dominance relation. There are several *prima facie* plausible reasons to believe this: the respect seems *gerrymandered*; it seems *evaluatively redundant* with the other respects in my example (such as the wellbeing levels of each individual); and there is a

⁷⁷ And perhaps there is for each respect *some* possible pair of things of which one is better in this respect without being worse in another respect. Perhaps, for example, there are some pairs of things of which only one has any good features and only one has any bad features.

complementary respect that counteracts this respect (i.e. total wellbeing of people with wellbeing 10). Each of these reasons corresponds to a potential solution to the application problem. I will discuss them in turn.

6.1 Ignoring Gerrymandered Respects

Nobody would normally compare two populations with respect to the total wellbeing of people with wellbeing 9 who are at least as well off as everyone else. This respect seems gerrymandered. Perhaps we can solve the application problem by ignoring such gerrymandered respects: it is sufficient that X is not worse than Y in any *non-gerrymandered* respect. If X is also better than Y in some non-gerrymandered respect, X is better overall.

The question is what exactly it means for a respect to be gerrymandered. What distinguishes gerrymandered respects from other respects? At least part of the reason why the respect of ‘total wellbeing of people with wellbeing 9 who are at least as well off as everyone else’ seems gerrymandered is that it is oddly specific. But the introduction of respects with a relatively high degree of specificity was essential for my defence of Weak Pseudo-Dominance. Even if an outcome is not worse for any person, it may be worse with respect to specific welfare goods of some person, such as their pleasure or even their enjoyment from playing music. Moreover, the application problem does not only arise from extremely specific respects. Total wellbeing of people with wellbeing 9 or lower is an unusual respect, but it isn’t highly specific. Given a simple axiology, overall betterness supervenes on sets of relatively few respects of roughly this degree of specificity. For example, on the total view, overall betterness supervenes on {total wellbeing of people with wellbeing 9 or lower, total wellbeing of people with wellbeing higher than 9}.

There is a feature that all respects that I used to demonstrate the application problem have in common: they refer to different individuals in the compared populations. For example, the people who have wellbeing 9 in population P are not identical with the people who have wellbeing 9 in population P+. This feature may contribute to the impression that the respects in my example are gerrymandered: it would be more natural to compare the populations in respects that refer to the same individuals in each population. However, this is not generally the case. There is nothing gerrymandered about comparing two populations with respect to the wellbeing of the worst-off people, even if the worst-off people are different individuals in the two populations. Similarly, it is natural to say that one action is better than another in the respect that it is better for one's future child even if the identity of the child depends on which action is performed. Surely, something is not even *pseudo-dominant* if it is worse for the worst-off people or for one's future child.

It seems difficult to find any general conditions that determine whether a respect is gerrymandered. There remains the possibility of postulating that there are *primitive facts* about which respects are gerrymandered. It is a common metaphysical view that there is a primitive distinction between natural and gerrymandered properties (see Lewis 1983). If one already accepts such a metaphysical distinction for other reasons, it may not constitute significant additional cost to postulate a similar distinction between different kinds of respects. Arguably it cannot be the same distinction, however. Or at least, it cannot be the distinction between properties that are *perfectly* natural and those that aren't: many of the respects in which we make evaluative comparisons have no chance of being perfectly natural. For example, there is nothing wrong about comparing two outcomes with respect to Bob's wellbeing, but Bob's wellbeing surely doesn't belong to any "élite minority of special properties", as Lewis (1983, 346) characterizes the perfectly natural properties. This is yet clearer for other respects in which we may legitimately compare things, such as Bob's level of stress in two different careers, or his athletic achievements at two different events.

These respects don't seem to come anywhere close to being perfectly natural. Hence, if we were to exclude certain respects based on their low degree of naturalness, the threshold for sufficient naturalness would need to be very low.

There is another, more severe problem with the attempt to solve the application problem by excluding gerrymandered respects: some of the respects that give rise to the problem don't seem gerrymandered once we consider them in certain practical contexts. Imagine, for example, that we compare two policies and it is to be expected that the choice between these policies will especially affect the total wellbeing of people with a wellbeing below 9. Imagine, moreover, that we have a reliable method of ascertaining how policies affect the total wellbeing of that subpopulation. In such an unusual scenario, it seems natural (or at least not particularly *unnatural*) to compare the two policies with respect to the total wellbeing of people with a wellbeing below 9. To decide between the two policies, we could start with a comparison in this respect and later on consider how the policies compare with respect to the total wellbeing of the remaining population, which is expected to be less affected. This approach seems reasonable and in its context the respect of total wellbeing of people with a wellbeing below 9 doesn't seem so strange or gerrymandered. The impression that it is gerrymandered seems to be at least partially explained by the fact that comparisons in this respect are just not useful for us. If we invented a technology with which we could easily detect how our options would affect total wellbeing of people with a wellbeing below 9, it would plausibly begin to seem natural to us to compare options in this respect.

6.2 Ignoring Evaluatively Redundant Respects

Another reason why the respects that create the application problem seem negligible is that they are *evaluatively redundant* with other respects that I had already considered. For

example, total wellbeing of people with wellbeing 9 seems evaluatively redundant with the set of respects corresponding to each individual's wellbeing. If we have already considered the latter, we don't additionally need to consider the former. It's not that there is something *intrinsically wrong* with the respect of total wellbeing of people with wellbeing 9 – in particular, it isn't gerrymandered in some problematic way. Rather, the respect just doesn't add anything evaluatively relevant *beyond* the other respects. So *if* we have already compared two populations with respect to every person's wellbeing as well as, perhaps, some other respects, such as equality, we can ignore this respect.

Part of the reason why the respect of total wellbeing of people with wellbeing 9 (let's call this respect "R") seems evaluatively redundant with the other respects is that it is not part of any *minimal* supervenience set that also includes the other respects: for any set of respects that includes R *and* the other respects (in particular: the wellbeing levels of each individual), if overall betterness supervenes on this set, it also supervenes on the subset that results from excluding R. R is, so to speak, an unnecessary addition to the supervenience set. But this is not the whole reason why R seems evaluatively redundant: R also doesn't ground any evaluatively relevant relation between the respects in the supervenience set or between these respects and the compared populations. In particular, R doesn't ground any *incommensurateness* or *non-separability* in the supervenience set.

Consider incommensurateness first. That population P+ is worse than population P with respect to total wellbeing of people with wellbeing 9 while being better in other respects, such as total wellbeing of people with wellbeing 10, means, technically, that there is some value conflict between P and P+. But this specific value conflict doesn't ground any incommensurateness between P and P+. After all, the case is designed such that there is (or could be) no incommensurateness between P and P+ at all. In particular, P+ is strictly better than P with respect to total wellbeing despite the above value conflict: we can just add up the total wellbeings of the two subpopulations to determine which population is

better with respect to total wellbeing. No complication arises that could ground incommensurateness.

Now consider non-separability. There plausibly are some “pattern goods”, such as equality and balance, that don’t allow us to consider certain respects (only) separately; if we want to infer how X and Y evaluatively compare overall from how they compare in these respects, we also need to consider how they jointly affect some pattern good. For example, on egalitarianism, the wellbeing levels of different individuals are not separable since how one individual wellbeing level affects equality depends on the wellbeing levels of other individuals. Consequently, the contribution of one individual wellbeing level to overall value depends on what wellbeing levels the other individuals have. More generally, we can say that one respect R1 is *non-separable* from another respect R2 if and only if how R1 affects the overall value difference between X and Y depends not only on the properties of X and Y that determine R1, but also on the properties of X and Y that determine R2. Pattern goods provide an explanation for the possibility of non-separability: how an R1-determinate affects overall value could depend on how R2 is determined because there could be an evaluatively relevant pattern that is itself a non-separable function of both R1-determinates and R2-determinates. Equality of wellbeing and balance between different welfare goods are plausible examples of such pattern goods. But total wellbeing of people with wellbeing 9 isn’t. In particular, it doesn’t ground non-separability between the wellbeing levels of different individuals. After all, total wellbeing of people with wellbeing 9 is a separable function of the wellbeing levels of people with wellbeing 9: the former is just the sum of the latter.

In sum, the respect R of total wellbeing of people with wellbeing 9 seems evaluatively redundant with a set S that contains all individual wellbeing levels because R is not part of any minimal supervenience set that contains S and because any incommensurateness or non-separability in S is not grounded in R. It seems plausible that

R is therefore negligible if we already consider the respects in S. More generally, it seems plausible that it suffices for overall betterness that X pseudo-dominates Y in a set of respects that is not only a supervenience set but also includes all respects that ground any instance of non-separability or incommensurateness within the set. Here is a slightly more sophisticated version of this principle.

Moderate Pseudo-Dominance: For all X and Y, X is better than Y overall if X pseudo-dominates Y in a set of respects S that meets the following conditions:

- (i) S is a supervenience set.
- (ii) Any subset of S that contains no respect in which X is better than Y is not a supervenience set.
- (iii) S contains all respects that ground non-separability between any subsets of S.
- (iv) S contains all respects that ground incommensurateness between X and Y in some respect in S.

Moderate Pseudo-Dominance is weaker than Strong Pseudo-Dominance since the sufficiency condition of the latter only requires that the set in which X pseudo-dominates Y meets condition (i), i.e. that it is a supervenience set. Moderate Pseudo-Dominance is stronger than Weak Pseudo-Dominance: conditions (i) to (iv) are met if X pseudo-dominates Y in the total set of all respects,⁷⁸ but – as we have seen – they can also be met by a subset of the total set: Moderate Pseudo-Dominance allows that X is worse than Y in

⁷⁸ That is obvious with respect to conditions (i), (iii), and (iv). With respect to condition (ii), the reasoning is a bit complicated, but here is the gist of it: if S is the total set, then S contains a *minimal* supervenience set in which X pseudo-dominates Y, and if X pseudo-dominates Y in a minimal supervenience set, *each* supervenience set contains a respect in which X is better than Y. The cases in which S does not meet condition (ii) are cases in which S does not contain any minimal supervenience set in which X pseudo-dominates Y. I will give an example later on.

some respects. Still, Moderate Pseudo-Dominance is barely less plausible than Weak Pseudo-Dominance. The main argument for Weak Pseudo-Dominance can be transformed into an argument for Moderate Pseudo-Dominance as follows. (It is supposed to hold for all X and Y, and all sets of respects S.)

- (1) Weak Incommensurateness Requirement: If X and Y are incommensurate in some respect R, then there is a non-empty set S1 that only contains respects in which X is better than Y, there is a non-empty set S2 that only contains respects in which Y is better than X, and the facts about how much X is better than Y in the respects in S1 and about how much Y is better than X in the respects in S2 at least partially ground the fact that X and Y are incommensurate in R.⁷⁹
- (2) If the Weak Incommensurateness Requirement holds, then, if X is not worse than Y in any respect in S, X is better than Y in some respect in S, and S includes all respects that ground incommensurateness between X and Y in some respect in S, X and Y are not incommensurate in any respect in S.
- (3) If X is not worse than Y in any respect in S, and X and Y are not incommensurate in any respect in S, then X is at least as good as Y in every respect in S.
- (4) Thus, if X is not worse than Y in any respect in S, X is better than Y in some respect in S, and S contains all respects that ground incommensurateness between X and Y in some respect in S, then X is at least as good as Y in every respect in S. (From (1), (2), and (3).)
- (5) Moderate Dominance: If X is at least as good as Y in every respect in S, X is better than Y in some respect in S, S is a supervenience set, any subset of S that contains

⁷⁹ The condition that X and Y must be commensurate in at least one respect is omitted to reduce the complexity of the argument. The condition is trivially met by any X and Y that meet the sufficiency condition of Moderate Pseudo-Dominance, which requires that X is better than Y in some respect.

no respect in which X is better than Y is not a supervenience set, and S contains all respects that ground non-separability among any of its subsets, then X is better than Y overall.

- (6) Moderate Pseudo-Dominance: If X is not worse than Y in any respect in S, X is better than Y in some respect in S, S contains all respects that ground incommensurateness between X and Y in some respect in S, S is a supervenience set, any subset of S that contains no respect in which X is better than Y is not a supervenience set, and S contains all respects that ground non-separability among any of its subsets, then X is better than Y overall. (From (4) and (5).)

Like the argument for Weak Pseudo-Dominance, this argument has two substantial premises: the Weak Incommensurateness Requirement (1) and a dominance principle (5). The conjunction of the Weak Incommensurateness Requirement and condition (iv) of Moderate Pseudo-Dominance shows that there is no incommensurateness in set S if X pseudo-dominates Y in S: if incommensurateness is (at least partially) grounded in value conflicts, there is no value conflict in S, and there are no respects outside of S that ground incommensurateness in S, then there is no incommensurateness in S. In the argument for Weak Pseudo-Dominance, it was not necessary to add condition (iv) because it is trivially met if S is the *total* set. Since, as the application problem shows, we need to relativize the pseudo-dominance relation to a subset, we also need to add condition (iv): otherwise some value conflicts between respects outside of S could ground incommensurateness in S.⁸⁰

The dominance principle in (5) – Moderate Dominance – can be presented analogously to Moderate Pseudo-Dominance.

⁸⁰ This is the case in the objections against pseudo-dominance principles that I discussed in section 3: these objections are based on cases in which X pseudo-dominates Y (only) in supervenience sets that don't include some respects that ground incommensurateness in these sets. Condition (iv) thereby ensures that Moderate Pseudo-Dominance, just like Weak Pseudo-Dominance, does not apply in these cases and is hence immune to the objections from the literature.

Moderate Dominance: For all X and Y, X is better than Y overall if X dominates Y in a set of respects S that meets the following conditions:

- (i) S is a supervenience set.
- (ii) Any subset of S that contains no respect in which X is better than Y is not a supervenience set.
- (iii) S contains all respects that ground non-separability between any subsets of S.

I take it that Moderate Dominance is highly plausible. In fact, it may seem unnecessarily weak. Why do we need conditions (ii) and (iii)?

The purpose of condition (ii) is to ensure that X dominates Y not only in S but also in every supervenience set that is a subset of S. In particular, the condition is supposed to ensure that S is not a union set of a supervenience set that only contains respects in which X and Y are equally good and a (non-supervenience) set of respects in which X is better than Y.⁸¹ In that case, X and Y would be equally good overall (given that the respects in the supervenience set are separable). For example, consider two populations containing just one individual, Amy, who has the same wellbeing level in each population. Clearly, these populations are equally good in every respect of a supervenience set, say {total wellbeing, equality}, and they are equally good overall. But Amy could have more pleasure in one population and more desire-satisfaction in the other population – the same wellbeing could be the result of different welfare goods. Thus, each population dominates the other in a supervenience set: one in {total wellbeing, equality, pleasure} and the other in {total

⁸¹ This condition is trivially fulfilled if S is a *minimal* supervenience set. But S won't be a minimal supervenience set if it contains any respects that ground non-separability between other respects in S, since the former supervene on the latter (see the subsequent discussion in the main text).

wellbeing, equality, desire-satisfaction}. Without condition (ii), we would get the absurd result that each of the populations is overall better than the other. With condition (ii), we avoid this result because each of these sets contains a supervenience set that only contains respects in which neither population is better than the other, i.e. {total wellbeing, equality}.

Now consider condition (iii) concerning non-separability. This condition is not trivially met by any set *S* that is a supervenience set. Supervenience sets don't need to include (and *minimal* supervenience sets *don't* include) respects that ground non-separability between other respects in the set because the former supervene on the latter. For example, equality of wellbeing supervenes on individual wellbeing levels, and balance between different welfare goods supervenes on these welfare goods. So it is possible that *X* dominates *Y* in a supervenience set but *X* is worse than *Y* with respect to certain pattern goods, such as equality and balance, that ground non-separability between the respects in the supervenience set. Condition (iii) rules out this possibility by requiring that any non-separability grounding respects are included in *S*: since *S* does not contain any respects in which *X* is worse than *Y*, this means that *X* is not worse than *Y* with respect to any pattern goods that ground non-separability. It is not clear whether this makes condition (iii) necessary because it could be that pattern goods cannot undermine overall betterness when *X* dominates *Y* – the dominance relation could always win out. Virtually all egalitarians, for example, hold that making one person better off while holding everyone else's wellbeing fixed makes an outcome overall better even if this change leads to more inequality – they are against “levelling-down”. But some authors have argued that the same does not hold for some other pattern goods. Chang (2002, 674) argues, for example, that increasing pleasure while holding all other welfare goods fixed could make a life not better

overall if the extra pleasure creates an imbalance with the other welfare goods.⁸² Condition (iii) ensures that Moderate Dominance is immune against such potential counterexamples.

In sum, conditions (i) to (iii) seem to rule out any reasonable concerns that one might have with dominance principles. Whether or not stronger dominance principles are defensible, Moderate Dominance is highly plausible. As the argument shows, the conjunction of Moderate Dominance and the intermediate conclusion (4), which rules out incommensurateness, imply Moderate Pseudo-Dominance.

I conclude that Moderate Pseudo-Dominance is barely less plausible than Weak Pseudo-Dominance and, unlike Weak Pseudo-Dominance, applicable. (An interesting application of it will be demonstrated in the next section.) Its only apparent downside is that it is a complicated principle. Besides the consequent lack of aesthetic appeal, this could matter for two reasons: Moderate Pseudo-Dominance is not *in itself* intuitively plausible, and it is difficult to apply. Both these concerns are mitigated (if not fully resolved) by the fact that Moderate Pseudo-Dominance is motivated by the simple intuition that Weak Pseudo-Dominance captures. Regarding plausibility: Moderate Pseudo-Dominance may not be intuitively compelling in itself, but Weak Pseudo-Dominance is, and it seems that if Weak Pseudo-Dominance is true, Moderate Pseudo-Dominance is true too. Regarding difficulty of application: Moderate Pseudo-Dominance will apply in some cases in which it is not obvious at all that it does apply – it is not always obvious whether certain respects ground incommensurateness, for example. But Moderate Pseudo-Dominance is fairly easy to apply in those cases in which it intuitively *seems* that Weak Pseudo-Dominance applies: those cases in which it intuitively *seems* that there is no respect at all in which X is worse than Y. The application problem shows that even in these cases there are *some* respects in which X is worse than Y. So we cannot apply Weak Pseudo-Dominance. But, as in the case

⁸² See also Chang (2016, 402) for some other examples.

of *total wellbeing of people with wellbeing 9*, it is easy to tell that a set does not need to include these respects in order to meet the conditions of Moderate Pseudo-Dominance. So we just apply Moderate Pseudo-Dominance instead.

6.3 Ignoring Outweighed Respects

The third reason why the respects that are responsible for the application problem seem negligible is that these respects can be matched with other respects, such that the *conjunction* of these respects favours the pseudo-dominant thing. In my example of the two populations P and P^+ , the respect of total wellbeing of people with wellbeing 9 can be matched with the respect of total wellbeing of people with wellbeing 10. Weak Pseudo-Dominance could not be applied since P^+ is worse than P with respect to the former. But P^+ is better than P with respect to the latter, and P^+ is *also* better than P with respect to the conjunction of both respects. After all, the total wellbeing of people with wellbeing 9 *or* 10 is higher in P^+ than in P . Since the conjunction of both respects favours P^+ over P , it seems that these respects cannot undermine the overall betterness of P^+ . If P^+ pseudo-dominates P in the set of all the remaining respects, this should be enough to make P^+ overall better than P .

More generally, the idea is that we can “outsource” those respects in which X is worse than Y – and that hence prevent the application of Weak Pseudo-Dominance – into another set along with (complementary) respects in which X is better than Y . If X is at least as good as Y with respect to the set of these outsourced respect, we can ignore this set and treat the set of all the remaining respects as if it was the total set. A simple way of making this idea precise is the following: X is overall better than Y if the total set of all respects can be partitioned into two sets, such that X pseudo-dominates Y in one set and X is at least

as good as Y with respect to (the conjunction of all respects in) the other set.⁸³ This principle looks initially plausible, but it needs to be amended. Note that, as it stands, the set in which X pseudo-dominates Y could be a set of a single respect in which X is better than Y. In many cases, pseudo-dominating in such a tiny set while being only equally as good with respect to the set of the remaining respects is not sufficient for overall betterness. That is because the single respect in the pseudo-dominance set could be (and probably is) evaluatively *redundant* with the respects in the extremely-encompassing set of the remaining respects. For example, the single respect could be Bob's pleasure. If X is better than Y in this respect, and X is as good as Y with respect to the set of all remaining respects, X won't be better than Y overall – X and Y will be equally good overall since the set of the remaining respects contains Bob's wellbeing relative to which Bob's pleasure is evaluatively redundant.

One possible way of amending the simple principle is to add requirements regarding the set in which X pseudo-dominates Y. In particular, one could add the condition that this set needs to be a supervenience set and – for the reasons discussed in the previous subsection – the condition that any of its subsets that contains no respect in which X is better than Y is not a supervenience set. Fortunately, there is a more elegant solution: we can just require that X is *better* than Y with respect to the set of the *outsourced* respects. Instead of making the pseudo-dominance set sufficiently encompassing such that the pseudo-dominance relation in this set ensures overall betterness, we ensure overall betterness through the set of the outsourced respects. Here, we don't need to add any further conditions about how encompassing the set is because it is guaranteed to have one crucial feature: it contains all respects in which X is worse than Y. That is has this feature follows

⁸³ For brevity, I will often just say that X is better than Y “with respect to a set” rather than that it is better “with respect to the conjunction of all respects in the set”.

from the condition that X pseudo-dominates Y with respect to the set of all remaining respects. This condition is equivalent with the conjunction of (i) the set of the *outsourced* respects contains *all* respects in which X is worse than Y and (ii) the set of the *outsourced* respects does *not* contain all respects in which X is better than Y. Since we want to recover overall betterness from the set of the outsourced respect, we can do away with (ii): it may as well contain *all* respects in which X is better than Y. What matters is simply that X is better than Y with respect to this set, and that this set contains all respects in which X is worse than Y.

Generalized Pseudo-Dominance: For all X and Y, if X is better than Y with respect to a set of respects that contains all respects in which X is worse than Y, then X is better than Y overall.

If X and Y meet the sufficiency condition of Weak Pseudo-Dominance, they also meet the sufficiency condition of Generalized Pseudo-Dominance. If X and Y meet the former condition, there is at least one respect in which X is better than Y and there is no respect in which X is worse than Y. Thus, one of the sets that contain all respects in which X is worse than Y is a set that contains nothing but one respect in which X is better than Y. And, trivially, X is better than Y with respect to such a set. But Generalized Pseudo-Dominance also applies in many cases in which Weak Pseudo-Dominance does not apply. For example, Generalized Pseudo-Dominance implies in the example that illustrates the application problem that population P⁺ is better than population P: there is a set that contains all respects in which P⁺ is worse than P with respect to which P⁺ is better than P. After all, we can outweigh any respect in which P⁺ is worse than P by a complementary respect in which P⁺ is better than P. Although not necessary, we could also add to this set all the other

respects in which P^+ is better than P , such as individual wellbeing levels, total wellbeing, average wellbeing, and so on. With respect to the whole set, P^+ is clearly better than P .

Generalized Pseudo-Dominance is an elegant principle. I find it also intuitively compelling – in fact, no less compelling than Weak Pseudo-Dominance. But one might be sceptical because Generalized Pseudo-Dominance seems not to rule out that X and Y are incommensurate in some respects. And indeed, it doesn't. Its sufficiency condition ensures overall betterness independently of whether there is incommensurateness in some respects. Here is the argument:

- (1) If X is better than Y with respect to a set that contains all respects in which X is worse than Y , then X is better than Y with respect to a set that contains not only all respects in which X is worse than Y but also all respects in which X is at least as good as Y .
- (2) Strong Incommensurateness Requirement: If X and Y are incommensurate in some respect R , then there is a non-empty set $S1$ that only contains respects in which X is better than Y , there is a non-empty set $S2$ that only contains respects in which Y is better than X , and the facts about how much X is better than Y in the respects in $S1$ and about how much Y is better than X in the respects in $S2$ fully ground the fact that X and Y are incommensurate in R .⁸⁴
- (3) If the Strong Incommensurateness Requirement holds, any set that contains all respects in which X is at least as good as or worse than Y contains all respects that ground incommensurateness between X and Y in any respect.

⁸⁴ As in the previous argument, the condition that X and Y must be commensurate in at least one respect is omitted to reduce the complexity of the argument. The condition is trivially met by any X and Y that meet the sufficiency condition of Generalized Pseudo-Dominance, which requires that X is better than Y in some respect (otherwise X couldn't be better than Y with respect to a set that contains all respects in which X is worse than Y).

- (4) Thus, if X is better than Y with respect to a set that contains all respects in which X is worse than Y, then X is better than Y with respect to a set that contains not only all respects in which X is worse than Y, but also all respects in which X is at least as good as Y, and all respects that ground incommensurateness between X and Y in any respect. (From (1), (2), and (3).)
- (5) If the Strong Incommensurateness Requirement holds, then, for any set S that contains all respects that ground incommensurateness between X and Y in any respect, if X is better than Y with respect to S, then X is also better than Y with respect to the union set of S and the set of all respects in which X and Y are incommensurate.
- (6) Thus, if X is better than Y with respect to a set that contains all respects in which X is worse than Y, then X is better than Y with respect to a set that contains not only all respects in which X is worse than Y, but also all respects in which X is at least as good as Y, all respects in which X and Y are incommensurate, and all respects that ground incommensurateness between X and Y in any respect. (From (2), (4) and (5).)
- (7) If X is better than Y with respect to a set that contains all respects in which X is worse than Y or at least as good as Y or incommensurate with Y (i.e. with respect to the total set of all respects), then X is better than Y overall.
- (8) Generalized Pseudo-Dominance: If X is better than Y with respect to a set that contains all respects in which X is worse than Y, then X is better than Y overall. (From (6) and (7).)

The general strategy of this argument is to gradually expand the set that contains all respects in which X is worse than Y until the set contains *all* respects, and to argue that it remains the case in each step that X is better than Y with respect to the expanding set. Premise (1)

adds to the set all respects in which X is *at least as good as* Y. The addition of such respects is innocuous; they cannot undermine X's betterness with respect to the set. In general, the only reason to believe that adding respects in which X is at least as good as Y to a set could undermine X's betterness with respect to the set is that the added respects are *non-separable* from the other respects in the set. But here, the set at issue already contains *all* respects in which X is worse than Y. Thus, it already contains all respects in which X is worse than Y that could ground non-separability. Such respects are, so to speak, already "priced in". For example, if X is worse than Y with respect to inequality or balance, these respects are already part of the initial set with respect to which X is better than Y.

The interesting part of the argument concerns the addition of respects in which X and Y are *incommensurate* to the set. To justify that X is still better than Y with respect to the set after the addition of these respects, the argument employs the Strong Incommensurateness Requirement (2) – which finally becomes relevant here. According to this requirement, any instance of incommensurateness is *fully* grounded in counteracting respects. Based on this principle, it is argued, first, that any set that contains all respects in which X is at least as good as or worse than Y contains all respects that *ground* incommensurateness between X and Y in any respect (3). This follows trivially from the Strong Incommensurateness Requirement. In a second step (5), it is then argued that adding respects in which X and Y are incommensurate to a set that already contains all respects that *ground* their incommensurateness cannot undermine X's betterness with respect to that set. I take it that this seems plausible in itself. But the claim is not trivially true since there might be things other than respects that ground incommensurateness, such as differing evaluative standards or differing ontological categories. For that reason, premise (5) is conditionalized on the Strong Incommensurateness Requirement, too. It ensures that *anything* – not only any *respect* – that grounds incommensurateness is a respect in which X is better or worse than Y. Given this, it seems highly plausible that if X is better than Y

with respect to a set that contains any respect – and hence *anything* – that grounds incommensurateness, X is also better than Y with respect to the set that additionally contains the respects in which X and Y are incommensurate.

We have thus expanded the set that contains all respects in which X is worse than Y to the set that additionally contains all respects in which X is at least as good as Y and all respects in which X and Y are incommensurate, i.e. the total set of *all* respects, without losing X's betterness with respect to the set (6). Trivially, if X is better than Y with respect to the total set – i.e. better with respect to the conjunction of *all* respects – X is better than Y overall (7). Thus, as Generalized Pseudo-Dominance states, X is better than Y overall if X is better than Y with respect to a set that contains all respects in which X is worse than Y (8).

I had commented in the context of Moderate Pseudo-Dominance that this principle will apply in some cases in which it is not obvious at all that it does apply. The same holds for Generalized Pseudo-Dominance. In these cases, the principle will be of little help. But, like Moderate Pseudo-Dominance, Generalized Pseudo-Dominance is fairly easy to apply in those cases in which it intuitively *seems* that Weak Pseudo-Dominance applies: those cases in which it intuitively *seems* that there is no respect at all in which X is worse than Y. The application problem shows that even in these cases there are *some* respects in which X is worse than Y. So we cannot apply Weak Pseudo-Dominance. But, as in the case of *total wellbeing of people with wellbeing 9*, it is usually easy to match these respects with other respects in which X is better than Y, such that X is better than Y with respect to their conjunction. We can then infer by Generalized Pseudo-Dominance that X is overall better than Y.

7. How Pseudo-Dominance Refutes Neutrality

In response to the application problem, I have developed two stronger versions of Weak Pseudo-Dominance: Moderate Pseudo-Dominance and Generalized Pseudo-Dominance. I will now put these two principles to work. I will show how they can be used to amend an argument from John Broome who uses a pseudo-dominance principle to argue against the neutrality view in population ethics.⁸⁵ According to that view, additions of happy people to a population that don't affect the wellbeing of other people (so-called "mere additions") are axiologically neutral: they make the world neither better nor worse.

Broome argues that the neutrality view is committed to a "greedy" sort of neutrality that can cancel out other values: if the existence of additional happy people is neutral by itself, as the neutrality view claims, the combination of additional happy people and other good or bad things will often be neutral too. For example, the combination of additional happy people and welfare decreases for others will often be neutral. In comparative terms, this means that the resulting population with the additional people and the decreased welfare will be neither better nor worse than the original population. But, according to Broome, this cannot be true since this population is worse in one respect and neither better nor worse in the other respect. With regard to a particular example, Broome writes the following:

These two options differ in two respects. First, one man is worse off in *C* than *A*. For instance, we may suppose his life is shorter in *C*; he lives eighty years in *A* and sixty in *C*, say. *C* is unequivocally worse than *A* in this respect. Second, *C* contains an extra person, whose wellbeing is in the neutral range. In this respect, *C* is neither better nor worse than *A*. That is the intuition we are working on: that adding a person is neutral. So *C* is worse than *A* in one respect, and neither better nor worse in the other respect. Intuitively, therefore, *C* must on balance be worse than *A*. (Broome 2005, 409)

⁸⁵ See Broome (2004, 170), (2005), and (2022). Broome dubs this view the "neutrality intuition" since he takes the view to be supported by a common intuition.

Strictly speaking, there are many more respects in which the compared populations differ than the two respects that Broome mentions. For example, the welfare decreases will go along with decreases in certain welfare goods and a lower average wellbeing while the addition of the happy person will increase total wellbeing and probably affect the degree of inequality in the population. Broome's argument is thus best understood as being implicitly relativized to a certain (supervenience) set of respects that fully determines the value difference between the two populations: within this set, A and C differ only in the two respects that Broome mentions, and since A is better than C with respect to one of them and not worse than C with respect to the other, A pseudo-dominates C in the set. Specifically, Frick (2017, 355) interprets Broome's argument as an application of the *decomposition principle*, which relativizes the pseudo-dominance relation to a set of respects that decomposes the total difference between the compared things, and Hedden (2024, 581) interprets Broome as relying on *individual-wise* pseudo-dominance, which relativizes the pseudo-dominance relation to the set of all individual wellbeing levels. Since these principles are incompatible with incommensurateness, as discussed in section 3, Frick and Hedden reject Broome's argument. More generally, conditional on the existence of incommensurateness, the mere fact that A pseudo-dominates C in a supervenience set is not sufficient for A's overall betterness – Strong Pseudo-Dominance is not true.

But we can salvage Broome's argument by applying Moderate Pseudo-Dominance or Generalized Pseudo-Dominance instead. In the particular case of populations A and C, it is unclear whether these principles can be applied since the case is underspecified – we would need to know more about how A and C compare in respects such as equality and average wellbeing. For example, if A is worse than C with respect to equality (as a figure with which Broome (2005, 408) illustrates the case suggests), Moderate Pseudo-Dominance is not applicable since equality is a pattern good that grounds the non-separability between individual wellbeing levels. Accordingly, the supervenience set in

which A pseudo-dominates C, which contains individual wellbeing levels, would then not contain all respects that ground non-separability between its elements and would thus fail to meet condition (iii) of Moderate Pseudo-Dominance. Moreover, advocates of the neutrality view could in principle claim that the existence of the additional happy person in C is good *in some respect*. After all, the neutrality view only states that the (mere) addition of a happy person makes a population neither better nor worse *overall*. Plausibly, this implies, as Broome claims, that a population that contains an additional happy person is not better than another population that doesn't contain this person *with respect to* the existence of that person. Still, advocates of the neutrality view could consistently claim that the population with the additional happy person is better with respect to *certain valuable features* of the additional person. Again, this would mean that there are some respects in which A is worse than C, which would make the application of Generalized Pseudo-Dominance and of Moderate Pseudo-Dominance questionable if not impossible.⁸⁶

To avoid this complication, I will use a case in which the neutrality view implies that the addition of happy people cancels out the value of welfare *increases* and argue that Moderate Pseudo-Dominance and Generalized Pseudo-Dominance imply that the population *with* the additional happy people and the welfare increases is better overall. Concretely, consider the following case, which is due to Francis (2022, 71).

	One hundred people	Ten billion different people
P	40	---
Q	40.01	40.01
R	40	60

⁸⁶ Unlike Weak Pseudo-Dominance, Generalized Pseudo-Dominance and Moderate Pseudo-Dominance allow that X is worse than Y in some respects. But, as I explained in the previous section, the more such respects there are, the less clear it becomes whether these principles apply. The strategy is therefore to rule out as many respects as possible in which X could be worse than Y.

The neutrality view entails that R is not better than P since R differs from P only in that it contains additional happy people. If R is not better than P, R is not better than Q or Q is not better than P. (Otherwise, if R was better than Q and Q was better than P, it would follow by the transitivity of ‘better than’ that R is better than P.) Moreover, R is clearly better than Q: the immense welfare gains for the ten billion people clearly outweigh the slight welfare losses for the one hundred people.⁸⁷ Thus, Q cannot be better than P if the neutrality view is true.

Now compare these two populations. Q is better than P with respect to the wellbeing of the one hundred people who are all better off in Q. Moreover, as the neutrality view agrees, the existence of the ten billion additional happy people in Q is at least not a respect in which Q is *worse* than P. So far, this is what all cases have in common in which additions of happy people are combined with welfare increases for others. But this case has further features that are important for my argument. The larger population, Q, is also perfectly equal and has higher average wellbeing than P. In fact, every person in Q is better off than every person in P. Hence, Q is (at least) not worse than P with respect to equality, average wellbeing, and the wellbeing of the worst off.

Q could still be worse than P in many respects. Francis’ case only tells us about the overall (lifetime) wellbeing of each individual, but it doesn’t tell us about the welfare values that are responsible for that wellbeing. If Q lacks some positive welfare values that are realized in P, or if Q includes some negative welfare values that are lacking in P, Q is worse than P in some respect. For example, the slightly higher wellbeing of the one hundred people in Q could be the result of much more pleasure and (somewhat) less desire

⁸⁷ As Francis (2022, 71) argues, that is so even if we give some priority to the wellbeing of the one hundred people since they are worse off overall. To resist the premise, one would basically need to give absolute priority to the worse off.

satisfaction, and the wellbeing of the additional people in Q could be the result of both pleasure and suffering. The first problem can be easily stipulated away: the one hundred people are better off in Q because the only difference to their lives in P is that they enjoy a bit more of one welfare value (say, pleasure) in Q. It is less obvious if it can be stipulated that the additional people in Q don't suffer any personal bads. After all, even amazing lives normally contain some suffering and every human life eventually ends in death.

However, even though such a life has presumably never been lived, it seems possible for a human life not to contain any personal bads.⁸⁸ We can imagine that the person would feel good at times but – whether due to natural circumstance or through advanced technology – never feels bad. The person has some – perhaps rather moderate – goals and preferences that are all fulfilled through the course of her life. Similarly, she forms some true and no false beliefs, and so on. The life would not even have to be particularly good overall: it could be pleasant but not very exciting and without great heights.⁸⁹ Now imagine that there is a point in time after which the life would not contain any goods anymore. All plans and projects have been completed at that point and there would be nothing enjoyable anymore. Exactly at that point the person falls into a dreamless sleep that lasts indefinitely.

Alternatively, if you are sceptical that this is even metaphysically possible, we can imagine that the person dies painlessly at this point in time. Given the unusual circumstances, even the person's death seems not to be bad for her. This intuitive judgement gets support from the most common accounts of the badness of death. The *deprivation account* (e.g. Nagel 1970; Feldman 1991) agrees that the death of that person is not bad for

⁸⁸ Compare with Thornley (2022, 395) who calls lives with only good and neutral components “straightforwardly better than blank”. He imagines them as containing only dreamless sleep followed by constant happiness.

⁸⁹ Accordingly, the neutrality view cannot be defended against this argument by restricting it to people whose lives don't exceed a very high quality. Broome (2004, 2005) defines the neutrality view broadly as to cover such views with a limited “neutral range”.

her in any way since her death does not deprive her of any goods. By assumption, there would have been no further goods in her life had she not died.⁹⁰ The *frustration account* (e.g. Williams 1973) agrees with that judgement since the person's death does not frustrate any of her projects or desires. That is because, at the time of her death, she does not have any desire to stay alive and all of her projects and plans have already been completed. Finally, the *epicurean account* (e.g. Rosenbaum 1986) on which death is *generally* not a personal bad obviously agrees with this verdict as well.

I conclude that the lives of the additional people in Q don't have to contain any personal bads. If they don't, their addition seems not bad in any respect: not only are their lives not bad *for them* in any respect, but they also don't negatively affect the wellbeing of anyone else, and their addition does not decrease average wellbeing, equality, or the wellbeing of the worst off. Moreover, given that the lives of the people who live in both Q and P are unambiguously better in Q without leading to any inequality, their welfare gains in Q are better in some respect and not worse in any respect. It seems, therefore, that P and Q can be specified in such a way that there is just no respect in which Q is worse than P while Q is better than P in many respects.⁹¹

But we know from the application problem that this cannot be true. Just as in my previous example, there are some respects that are tailored to some good features of P or to some bad features of Q. For example, Q is worse than P with respect to total wellbeing of people with wellbeing 40 (since there are people with this wellbeing only in P). This means

⁹⁰ Arguably, it is not even necessary to assume that the deaths of the additional people don't deprive them of any goods. On a natural interpretation, the deprivation account considers someone's death only comparatively bad relative to possible worlds in which the person does exist but dies at a later time. Here, however, the relevant comparison is with a world in which the additional people do not exist at all. Relative to such a world, the deaths of additional people do not seem to be bad on the deprivation account *even if* they deprive the additional people of many goods.

⁹¹ It could be objected that the added people suffer *comparative harms* since they are less well off than in R (see Roberts 2022). This proposal has some plausibility if we assume that both Q and R were options for some agent who could choose between them. But we don't need to make that assumption: Q and R are just two possible populations, not options of any agent.

that we cannot apply *Weak* Pseudo-Dominance. But we can apply Generalized Pseudo-Dominance and Moderate Pseudo-Dominance. Generalized Pseudo-Dominance applies because every tailored respect in which P is better than Q can be matched with a corresponding respect in which Q is better than P, such that their conjunction favours Q. For example, Q is better than P with respect to the conjunction of the respects in the set {total wellbeing of people with wellbeing 40, total wellbeing of people with wellbeing 40.01}. More generally, Q is clearly better than P with respect to *some* set that contains all respects in which Q is worse than P – this set could contain total wellbeing, average wellbeing, wellbeing of the worst off, and all the other respects in which Q is better than P. Alternatively, we can apply Moderate Pseudo-Dominance since the tailored respects in which Q is worse than P are evaluatively redundant with the set of the other respects in which Q pseudo-dominates P. In particular, the tailored respects ground neither incommensurateness nor non-separability among the other respects. Thus, if Generalized Pseudo-Dominance or Moderate Pseudo-Dominance are true, the neutrality view is false.

8. Conclusion

Pseudo-dominance principles are commonly rejected because their antecedents don't rule out that the compared things are incommensurate in some respects. I have argued that if we take seriously the condition of the simple pseudo-dominance principle (i.e. Weak Pseudo-Dominance) that the pseudo-dominant thing is not worse in any respect, the problems with incommensurateness can be avoided, but at the cost of inapplicability. The difficulty is to find a pseudo-dominance principle that can handle incommensurateness without becoming inapplicable. I have argued that two specific principles overcome this difficulty: Moderate Pseudo-Dominance and Generalized Pseudo-Dominance. The former principle is closest to the original Weak Pseudo-Dominance; its antecedent rules out incommensurateness by

ensuring that the pseudo-dominant thing is only worse in evaluatively redundant respects. Generalized Pseudo-Dominance departs significantly from Weak Pseudo-Dominance and other pseudo-dominance principles. Its antecedent allows worseness in all kinds of respects and thus does not rule out that the compared things are incommensurate in some respects. I have argued that the principle is nevertheless compelling and that it can be justified by an incommensurateness requirement that is similar to the requirement that justifies Weak and Moderate Pseudo-Dominance. Both Moderate and Generalized Pseudo-Dominance can do important work. I have demonstrated one important implication: they imply that the neutrality view in population ethics is false.

References

- Audi, Paul. 2012. "Grounding: Toward a Theory of the In-Virtue-Of Relation." *The Journal of Philosophy* 109: 685–711.
- Bader, Ralf M. 2018. "Stochastic Dominance and Opaque Sweetening." *Australasian Journal of Philosophy* 96 (3): 498–507.
- Broome, John. 2004. *Weighing Lives*. Oxford: Oxford University Press.
- . 2005. "Should We Value Population?" *Journal of Political Philosophy* 13 (4): 399–413.
- . 2019. "Lessons from Economics." In *The Oxford Handbook of Ethics and Economics*, edited by Mark D. White.
<https://doi.org/10.1093/oxfordhb/9780198793991.001.0001>.
- . 2022. "Climate Change and Population Ethics." In *The Oxford Handbook of Population Ethics*, edited by Gustaf Arrhenius et al.
<https://doi.org/10.1093/oxfordhb/9780190907686.001.0001>.
- Chang, Ruth. 1997. "Introduction." In *Incommensurability, Incomparability, and Practical Reason*, edited by Ruth Chang. Cambridge: Harvard University Press.
- . 2002. "The Possibility of Parity." *Ethics* 112: 659–688.
- . 2016. "Parity: An Intuitive Case." *Ratio* 29 (4): 395–411.
<https://doi.org/10.1111/rati.12148>.

- Cusbert, John, and Robyn Kath. 2019. "A consequentialist account of Narveson's dictum." *Philosophical Studies* 176 (7): 1693–1709.
- Dorr, Cian, Jacob M. Nebel, and Jake Zuehl. 2023. "The case for comparability." *Noûs* 57 (2): 414–453.
- Feldman, Fred. 1991. "Some Puzzles About the Evil of Death." *Philosophical Review* 100 (2): 205–227.
- Fleurbaey, Marc. 2015. "Equality Versus Priority: How Relevant Is the Distinction?" *Economics and Philosophy* 31: 203–217.
- Francis, Tomi. 2022. *Totalism*. DPhil Thesis, University of Oxford.
- , and Johan E. Gustafsson. 2025. "Stochastic Dominance for Incomplete Preferences." *Analysis*: anaf046. <https://doi.org/10.1093/analys/anaf046>.
- Frick, Johann. 2017. "On the survival of humanity." *Canadian Journal of Philosophy* 47 (2-3): 344–367.
- Griffin, James. 1986. *Well-Being: Its Meaning, Measurement and Importance*. Oxford: Clarendon Press.
- Hare, Caspar. 2010. "Take the Sugar." *Analysis* 70 (2): 237–247.
- . 2013. *The Limits of Kindness*. Oxford: Oxford University Press.
- Hedden, Brian. 2024. "Parity and pareto." *Philosophy and Phenomenological Research* 109 (2): 575–592.
- , and Daniel Muñoz. 2024. "Dimensions of value." *Noûs* 58 (2): 291–305.
- , and Jacob M. Nebel. 2024. "Multidimensional concepts and disparate scale types." *Philosophical Review* 133 (3): 265–308.
- Lederman, Harvey. 2023. "Incompleteness, Independence, and Negative Dominance." arXiv:2311.08471. <https://arxiv.org/abs/2311.08471>.
- Lewis, David. 1983. "New work for a theory of universals." *Australasian journal of philosophy* 61 (4): 343–377.
- Mill, John Stuart. (1863) 1998. *Utilitarianism*. Edited by Roger Crisp. Oxford: Oxford University Press.
- Nagel, Thomas. 1970. "Death." *Noûs* 4 (1): 73–80.
- Narveson, Jan. 1973. "Moral problems of population." *The Monist* 57 (1): 62–86.
- Nebel, Jacob M. 2015. *Incommensurability in Population Ethics*. Unpublished BPhil dissertation.

- . 2019. "Asymmetries in the Value of Existence." *Philosophical Perspectives* 33: 126–145.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- . 1986. "Overpopulation and the Quality of Life." In *Applied Ethics*, edited by Peter Singer, 145–164. Oxford: Oxford University Press.
- . 2016. "Can we avoid the repugnant conclusion?" *Theoria* 82 (2): 110–127.
- Rabinowicz, Wlodek. 2009. "Broome and the Intuition of Neutrality." *Philosophical Issues* 19: 389–411.
- Raz, Joseph. 1986. *The Morality of Freedom*. Oxford: Clarendon Press.
- Roberts, Melinda. 2022. "The Person-Based Intuition and the Better Chance Puzzle." In *The Oxford Handbook of Population Ethics*, edited by Gustaf Arrhenius et al. <https://doi.org/10.1093/oxfordhb/9780190907686.001.0001>.
- Rosen, Gideon. 2010. "Metaphysical Dependence: Grounding and Reduction." In *Modality: Metaphysics, Logic, Epistemology*, edited by B. Hale and A. Hoffman, 109–35. Oxford: Oxford University Press.
- Rosenbaum, Stephen E. 1986. "How to Be Dead and Not Care: A Defense of Epicurus." *American Philosophical Quarterly* 23 (2): 217–225.
- Sartre, Jean-Paul. 1975. "Existentialism is a Humanism." In *Existentialism from Dostoevsky to Sartre*, edited by Walter Kaufmann. New York: Meridian-New American.
- Steele, Katie. 2022. "Incommensurability That Can(Not) Be Ignored." In *Value Incommensurability*, edited by Henrik Andersson and Anders Herlitz, 231–246. New York: Routledge.
- Thornley, Elliott. 2022. "Critical levels, critical ranges, and imprecise exchange rates in population axiology." *Journal of Ethics & Social Philosophy* 22: 382.
- Williams, Bernard. 1973. "The Makropulos case: Reflections on the tedium of immortality." In *Problems of the Self: Philosophical Papers 1956–1972*, 82–100. Cambridge: Cambridge University Press.
- Wilson, Jessica. 2023. "Determinables and Determinates." *The Stanford Encyclopedia of Philosophy* (Spring 2023 Edition), edited by Edward N. Zalta and Uri Nodelman. <https://plato.stanford.edu/archives/spr2023/entries/determinate-determinables/>.

Chapter 3: Neutrality about Making Happy People and the Badness of Human Extinction

Abstract

Human extinction in the near term would be terribly bad given optimistic assumptions about how humanity would otherwise develop in the future. This intuitive datum be easily explained by the view that a world is better if it contains more happy people: human extinction would mean the loss of countless good future lives. But many philosophers are attracted instead to the view that it is generally a *neutral* matter whether there are additional happy people. It is generally believed that the badness of human extinction can be explained without abandoning this neutrality view. After all, alternative explanations could draw on the harms to present people or the loss of non-welfarist values. This paper argues that this belief is mistaken: if the neutrality view is true, human extinction would not be bad overall. This is so even if we are extremely optimistic about the future of humanity.

1. Introduction

Given optimistic assumptions about the future development of humanity, it seems that human extinction in the near future would be terribly bad. If humanity would otherwise flourish while living in harmony with other species and the environment, humanity's continued existence would be much better than its extinction. What explains the value of humanity's survival and, correspondingly, the badness of human extinction under such optimistic assumptions?

A straightforward answer is that there would be very many individuals with good lives. All these good lives, and all the welfare values that make them good, would be lost if humanity went extinct. If we expect that huge numbers of people will exist in the future in case humanity avoids near-term extinction, as many have argued we *should* expect, near-term human extinction would therefore destroy an enormous amount of expected value (e.g. Bostrom 2013; Greaves & MacAskill 2025).

While this is a straightforward answer, it conflicts with a widespread intuition according to which it is usually a *neutral* matter whether or not there are additional people with good lives; bringing such “happy people” into existence does not make the world better *merely because* they would have good lives. As Jan Narveson (1973, 80) famously put it, “we are in favour of making people happy, but neutral about making happy people”. Friends of this *neutrality view* have therefore sought alternative explanations for the badness of human extinction.⁹² It has been argued, for example, that there are non-welfarist values that are necessarily connected to humanity or that humanity has final value in itself (e.g. Frick 2017; Henning 2022, 143). Moreover, standard examples of human extinction, such as extinction from nuclear war or an asteroid strike, would severely harm present people. It is indeed hard to imagine how humanity could soon go extinct without severely harming at least a large fraction of the last human generation.⁹³ For that reason alone, it seems, near-term human extinction would be one of the worst events, and probably *the* worst event, in human history.

Given optimism about humanity’s future, it therefore seems that we can all agree that human extinction would be terribly bad. This *convergence view* seems initially very

⁹² As Broome (2022, 396) notes, “many people, including many who are attracted by the intuition of neutrality, have a contrary intuition about extinction. They think that extinction would be very bad indeed.”

⁹³ Besides the obvious physical harm and the harm of premature death associated with most extinction events, human extinction could also harm people by undermining the value of their future-oriented projects (see Scheffler 2018).

plausible, and it has been endorsed by philosophers who do not endorse the neutrality view themselves. Toby Ord (2020), for example, takes the loss of good human lives to (at least partially) explain the badness of human extinction but argues that “there remain explanations based on the great things we could achieve in the future [...] and on areas other than our future: on our past, our character, our cosmic significance and on the losses to the present generation. People who hold person-affecting accounts of the value of the wellbeing of future generations could well be open to these other sources of reasons why extinction would be bad” (ibid. 124).⁹⁴ This would be an extremely important finding: we would not have to decide on population ethics, which is riddled with paradoxes, to conclude that human extinction would be terrible.

Unfortunately, the convergence view is false. If the neutrality view is true, human extinction would not be bad overall, even if we are very optimistic about the future of humanity. My argument for this claim is based on Broome’s (2004, 170) “greediness” argument which shows that if the addition of happy people is neither good nor bad, the *combination* of additional happy people *and* welfare increases or decreases for others is often neither good nor bad either. It has been noted before that this could mean that, on the neutrality view, human extinction is not bad despite its detrimental effects on present people. According to Steele (2025, 61), “neutrality allows that (1) [peace] and (3) [a nuclear war that kills 100%] are incommensurable”. Relatedly, Mogensen (2025, 100) notes that “the neutrality associated with the addition of worthwhile lives to the population threatens to swallow up and neutralize those other values [that speak in favour of a continued human presence], forcing us toward the conclusion that extinction would not be worse than continued human survival.” In this paper, I will argue more strongly that the neutrality view

⁹⁴ MacAskill (2022, 171–172) makes a similar claim. See also Knutzen (2023) who develops a view of the badness of human extinction that “may offer a way of taking a pro-humanity stand [i.e. the view ‘that it would be good for humanity to continue and bad for it to go extinct’] while accepting axiological neutrality” (ibid., 2).

needs to accept this type of verdict on pain of absurdity. It needs to accept that near-term human extinction would not be bad even under *very* optimistic assumptions about the future of humanity, *whatever* may be valuable about humanity's survival in such a case.⁹⁵

I will *not* argue that we don't have any *reason* – moral or otherwise – to prevent human extinction if the neutrality view holds.⁹⁶ Obviously, this normative claim follows from my axiological conclusion according to consequentialism. But as far as my argument is concerned, there could be non-consequentialist reasons to prevent human extinction that are independent of its badness. In fact, this is a common view (e.g. Finneron-Burns 2017, Lenman 2002, Scheffler 2018). Still, my argument won't be without a normative upshot for anyone who believes that there is at least *also* a moral reason to prevent bad outcomes. I will elaborate on this normative upshot in the conclusion. There, I will also briefly discuss how my argument can be extended to common variations of the neutrality view, such as to the view that the addition of happy people is neutral *if* their wellbeing falls into a certain range, and to the view that while adding happy people is never good, it can be *bad* if there are alternatives in which these people are better off. Until then, my argument will assume the simple neutrality view, as stated above, and it will be purely axiological.

I will begin by introducing a set of optimistic assumptions about humanity's future that jointly seem to guarantee that humanity's survival would be better than its extinction. For concreteness, I will introduce a survival outcome ("Survival") and an extinction outcome ("Extinction") that meet these assumptions. I will then argue that, against the intuitive comparison of these outcomes, Survival is *not* better than Extinction if the neutrality view

⁹⁵ As I will elaborate later, this includes non-welfarist values that might favour humanity's survival. The extension of my argument to non-welfarist values is an important contribution of this paper to the literature. But my argument also differs from the arguments in the literature in other important respects. For example, unlike Steele's argument, it does not assume any particular axiology for fixed-population cases, and it establishes a case of human survival that seems robustly good.

⁹⁶ A structurally analogous argument for this and similar normative claims could be made, but it would be far more controversial since it would require that the relevant normative relation, such as 'ought to _ rather than _' or 'has stronger reason to _ than to _', is transitive, which many reject (see Thomas 2023).

is true. I will make this argument in two steps. First, I will argue that the neutrality view cannot hold that Survival is better than Extinction in virtue of *welfarist* values. In particular, the harms to the present population in Extinction do not allow the neutrality view to hold that Survival is better than Extinction. To show this, I will initially assume that there are no non-welfarist values. In the second step of my argument, I will drop this assumption and argue that even if non-welfarist values favour Survival over Extinction, Survival is not overall better than Extinction if the neutrality view is true.

2. Optimism about the Future of Humanity

It is not obvious that humanity's near-term extinction would be bad because it is uncertain how humanity would otherwise develop in the future. If humanity will cause environmental collapse along with the extinction of most other species and (as a consequence) future human lives will be not worth living, it would be better if humanity went extinct sooner rather than later. If, however, we are sufficiently optimistic about both humanity's impact on the environment and its own flourishing, near-term human extinction seems clearly bad. Consider the following (perhaps too) optimistic assumptions about the future of humanity conditional on the avoidance of its near-term extinction.

- (1) All present people will be better off than they would be if humanity went extinct.
- (2) The average wellbeing of future people will be at least as high as the average wellbeing of the last few generations.⁹⁷
- (3) The worst-off future people will be no worse off than the worst-off people of the last few generations. More generally, relatively badly-off future people will be no

⁹⁷ Here and elsewhere I mean people's *lifetime* wellbeing. Since the lifetime wellbeing of the present generation has not been fully realized yet, I am referring in (1) to the lifetime wellbeing of the last few generations. We could add that the average wellbeing of future people will be at least as high as the *expected* average wellbeing of the present generation.

worse off than the people of the last few generations who were relatively badly-off.⁹⁸

(4) All future people (including the worst off) will have lives that are worth living.⁹⁹

(5) There will be many future generations under these welfare conditions.

(6) Other species and the environment will not be significantly harmed as a consequence of humanity's continued existence.

(7) Humanity will develop in such a way that non-welfarist values, if there are any, favour humanity's survival over its extinction.

Assumptions (1) to (4) determine how well off humans will be at the least in case of humanity's survival. The first assumption addresses the welfare of presently existing people. It generalizes what is to be expected by stating that *all* present people will be better off if humanity survives.

Assumptions (2) and (3) jointly guarantee that future generations will do at least as well as the last few generations: not only will the average wellbeing be at least as high; there will also be no change at the cost of those who are badly off compared to the rest of the population. Assumption (4) adds that future lives will be worth living. The addition of this assumption is important because some parts of the last few generations might have lives that are not worth living. Plausibly, this is indeed the case for some unfortunate individuals. Whatever the fraction of the last few generations with lives that are not worth living, assumption (4) implies that this fraction of the worst-off future people will be better off to the point that their lives are worth living. Together, these four assumptions seem to make

⁹⁸ Assumption (3) could be precisified as follows, in which case it implies (2): For each percentile P, the wellbeing corresponding to P of the worst-off people in the future is at least as high as the wellbeing corresponding to P of the worst-off people of the last few generations. (By quantifying over *each* percentile, it is guaranteed that the average wellbeing will be at least as high in the future.)

⁹⁹ We could add: and *most* future lives need to be *well* worth living. This will be the case in the specific survival outcome that I will introduce shortly.

humanity’s survival a good deal for humanity itself. That is especially so if humanity will not soon go extinct due a different event anyway, so that there will be many more future generations under these welfare conditions, as assumption (5) adds. Assumption (6) ensures that humanity’s survival will not be at the cost of other species or the environment: they will not be significantly harmed. Finally, assumption (7) adds that if the value of outcomes is affected by any non-welfarist considerations, these will overall favour humanity’s survival as well.

Some of these assumptions are probably too optimistic. In particular, it seems too optimistic that *all* future human lives will be worth living and that there will be *no* significant harms to the environment. I make these assumptions to stipulate a case of human survival that is both simple and clearly good. I take the assumptions to simplify, in a yet more optimistic way, actual optimistic views on the future of humanity that ground the belief that human extinction would be bad. But whether that is true is ultimately unimportant. Whatever one may think about the actual prospects of humanity, it seems clear that human extinction *would* be bad *if* these assumptions were accurate.¹⁰⁰ Consider, for concreteness, the following pair of an extinction outcome and a survival outcome, which meets the above assumptions.

Extinction	(2	/	/	/	/	/	/	...	/	/)
Survival	(5	6	7	8	9	9	9	...	9	2)

Each number represents the average wellbeing of one generation: the first number in each outcome represents the average wellbeing of the present population (which is treated as one generation), and the subsequent numbers represent the average wellbeing of the future

¹⁰⁰ Plausibly, human extinction would be bad even if some of these assumptions are not accurate. Temkin (2008, 208) and Beckstead (2013, 86) hold, for example, that human extinction would be bad even if it made present people overall *better* off. However, this is contentious. By contrast, I take my assumptions (1) to (7) to establish a case of human extinction that everyone would intuitively agree to be bad.

generations, which are assumed to be about as large as the present population. In Survival, the present population reaches an average wellbeing of 5, which represents good lives that are well worth living, and which is at least somewhat higher than the average wellbeing of the last few generations.¹⁰¹ The wellbeing of the next generations is higher (due to advances in agriculture and medicine, let's say) and keeps increasing until it reaches level 9. Each of these increases is supposed to be a significant increase in wellbeing, corresponding (for example) to several additional years of life with a high quality. Wellbeing 9 corresponds, accordingly, to very good lives – perhaps as good as the lives of some of the best-off present people. This high wellbeing level is then held constant for all subsequent generations until the last generation which is much worse off due to an extinction event that occurs during that generation. Their lives are still worth living but not by a very large margin: they are not *barely* worth living, but another global catastrophe during their lives would have made their lives not worth living. In Extinction, the present population has that same low wellbeing level due to a comparable extinction event. Since this extinction event occurs during the present generation, there are no future generations in Extinction.

The pair of Survival and Extinction clearly meets (and goes beyond) some of the optimistic assumptions that I made above. The present population is *much* better off in Survival than in Extinction (1), and there will be many future generations in Survival (5) with an average wellbeing that is *much* higher than the average wellbeing of the last few generations (2). To ensure that assumptions (3) and (4) are met as well, I will assume that there is perfect intragenerational equality in Survival, so that the wellbeing of each individual is as high as the average wellbeing of its generation. Finally, it is assumed in line with (6) that Survival does not include significant harms to other species or the

¹⁰¹ If you think that the average wellbeing of present people is so far lower than in previous generations, assume that there will be important technological advances in the near future that will benefit the present population.

environment, and in line with (7) that Survival is better than Extinction in non-welfarist respects if there are any non-welfarist values. With regard to (6), we can imagine, for example, that factory farming will be abolished soon and that all technologies that are harmful to the environment quickly get replaced by new technologies that are environmentally friendly. What (7) means concretely for how humanity develops in Survival depends on what non-welfarist values there actually are. It could mean, for example, that future generations will make certain valuable achievements. Or it could not require anything on top of the other assumptions because there are no non-welfarist values or because non-welfarist values necessarily favour humanity's survival anyway. I will discuss various potential non-welfarist values later on. But whatever non-welfarist values there are, if there are any, they overall favour Survival. Given all of this, it seems that it would be much better if Survival occurred than if Extinction occurred.

But the neutrality view cannot agree with this verdict. In the next section, I will argue for this claim on the assumption that there are only welfarist values: everything that affects the value of an outcome supervenes on the outcome's welfare distribution. Accordingly, I will simply compare the welfare distributions of Survival and Extinction.¹⁰²

3. No Badness despite Personal Harms

If the existence of additional good lives is in itself neither good nor bad but axiologically neutral, as the neutrality view claims, one would expect that the *combination* of additional

¹⁰² The welfare distributions given above do not include all welfare subjects who exist in Survival and Extinction: there are also non-human animals and past people. But I have assumed that humanity's survival will not significantly harm members of other species. I will also assume that the welfare of past people will not be affected by whether Extinction or Survival occurs. This is trivially the case if, in general, there are no posthumous harms; if there are posthumous harms, my assumption might still hold contingently; it is also compatible with the view that we would *wrong* past people if we *allowed* Extinction to occur (see Kaczmarek & Beard 2020). I make this assumption mostly for simplicity; my argument could be restated to cover potential welfare effects on past people without losing much force.

good lives *and* welfare increases for others is overall good. Accordingly, one would also expect that Survival is (much) better than Extinction if the neutrality view is true: all the additional people in Survival have good lives and the present population is much better off than in Extinction. In other words, one would expect that the neutrality view can account for the badness of Extinction in virtue of the harms to presently existing people.

An argument from John Broome (e.g. 2004, 2022) calls this expectation into doubt. According to his “greediness” argument, the combination of additional good lives and welfare increases or decreases will often be neither good nor bad if the neutrality view is true.¹⁰³ Importantly, Broome does not argue that this is *always* the case: some combinations of additional good lives and welfare increases may be good even on the neutrality view. So we cannot simply infer from his argument that Survival is not better than Extinction if the neutrality view is true. But we can implement his general argumentative strategy to argue for this specific claim.

Consider the following outcome in comparison to Survival and Extinction.

Alternative	(2	10	10	10	10	10	10	...	10	10)
Survival	(5	6	7	8	9	9	9	...	9	2)
Extinction	(2	/	/	/	/	/	/	...	/	/)

Alternative differs from Extinction in that it contains many additional people with very good lives. According to the neutrality view, this difference is a neutral matter: the mere fact that the additional people would have good lives does not make Alternative better. Alternative could still be better than Extinction in non-welfarist respects. But on my current assumption, there are no non-welfarist values. Given that assumption, the neutrality view implies that Alternative is not better than Extinction. Due to the transitivity of ‘better than’,

¹⁰³ While it has been controversially discussed whether this implication makes the neutrality view implausible, it is largely accepted that this is an implication of the view (e.g. Rabinowicz 2009, Frick 2017, Francis 2022).

this means that Alternative is not better than Survival *or* Survival is not better than Extinction: if both of the latter were the case, that is, if it was the case that Alternative is better than Survival and Survival is better than Extinction, it would follow by transitivity that Alternative is better than Extinction, contradicting the implication of the neutrality view. Thus, if the neutrality view holds, Survival is not better than Extinction *if* Alternative is better than Survival.

It remains to be seen whether Alternative is better than Survival. The people in Alternative are exactly the same as the people in Survival and, as in Survival, there is perfect equality within each generation. Their wellbeing levels are shown in the same order as in Survival, so that numbers in one column refer to the same groups of people. For the time being, we can imagine that in Alternative these groups of people form consecutive future generations just as in Survival. Accordingly, the differences between Alternative and Survival are simply that the present generation is better off in Survival while all future generations are better off in Alternative. The latter difference is clearly more significant. The higher wellbeing for the last generation in Alternative outweighs by itself the higher wellbeing for the present generation in Survival: the difference for the last generation is that between wellbeing 2 and wellbeing 10 while the difference for the present generation is only that between 2 and 5. On top of that, all the other future generations are significantly better off in Alternative. Hence, Alternative is clearly better than Survival. It follows by the above argument that Survival is not better than Extinction if the neutrality view is true.

We can also give a more theoretical argument for the premise that Alternative is better than Survival: it follows from the principles of *Weak Impartiality* and *Same-Person Pareto*. Weak Impartiality states that two outcomes that contain exactly the same people and only differ in the wellbeing of these people are equally good if each wellbeing level is

instantiated by the same number of people in both outcomes.¹⁰⁴ According to Weak Impartiality, Alternative is as good as the following outcome.

Alternative* (10 10 10 10 10 10 10 ... 10 2)

Now compare Alternative* with Survival.

Survival (5 6 7 8 9 9 9 ... 9 2)

For this comparison, we can apply Same-Person Pareto which states that, for all pairs of outcomes that contain exactly the same people and only differ in the wellbeing of these people, if in one of the two outcomes everyone is at least as well off as in the other outcome and some people are better off than in the other outcome, this outcome is better than the other outcome. Survival and Alternative* have exactly the same people, the last generation is equally well off in both outcomes and all other generations are better off in Alternative*. Hence, by Same-Person Pareto, Alternative* is better than Survival. Since, by Weak Impartiality, Alternative is as good as Alternative*, it follows that Alternative is better than Survival.

This argument could be restated with even weaker premises. In particular, we could abandon the (very plausible!) principle of Weak Impartiality and replace it by a principle that only rules out *extreme* partiality toward present people. Unlike Weak Impartiality, such a principle would allow that Alternative* is not as good as but *better* than Alternative. We could still use this weaker principle as long as it rules out that the value difference between Alternative* and Alternative is *as large* as the value difference between Alternative* and

¹⁰⁴ This is a logically weaker version of a principle known as *Impartiality* (Broome 2004, 135) or *Anonymity* (Francis 2021) which states that two outcomes are equally good if each wellbeing level is instantiated by the same number of people in both outcomes, *whether or not* the same people exist in both outcomes. This principle is popular but controversial since it contradicts person-affecting views by implying that the identities of the people in an outcome are axiologically irrelevant. Weak Impartiality, by contrast, only applies to the comparison of outcomes that contain the same individuals, and hence has by itself no implication for populations ethics.

Survival. And that the value difference between Alternative* and Alternative is not *that* large seems close to indubitable: whatever the additional value from swapping the wellbeing levels between the present and the last generation may be, it is surely smaller than the value that is lost from making all generations except the last one significantly worse off in moving from Alternative* to Survival. In moving from Alternative to Alternative* and then to Survival, we therefore end up with a worse outcome whether or not the move from Alternative to Alternative* is an improvement: if it is, it is clearly outweighed by the worsening in moving from Alternative* to Survival.

The premise that Alternative is better than Survival is thus reaffirmed. As before, we can now infer that Survival is not better than Extinction if the neutrality view holds, given the assumption that there are no non-welfarist values. Before complicating matters by dropping this assumption, it is worth noting that the argument can be extended to versions of Survival with even better welfare distributions. Consider, for example, the following improved version of Survival, in which the wellbeing of each generation except the last one is increased by 10.

Survival+ (15 16 17 18 19 19 19 ... 19 2)

Basically the same argument as before can be made. We only need to introduce a correspondingly improved version of Alternative.

Alternative+ (2 20 20 20 20 20 20 ... 20 20)

Alternative+ is better than Survival+. By the neutrality view, Alternative+ is not better than Extinction. Thus, by transitivity, Survival+ is not better than Extinction if the neutrality view holds.

To argue for the premise that Alternative+ is better than Survival+ we can give again the argument from Weak Impartiality and Same-Person Pareto. This is not possible for *all* improved versions of Survival. Consider, for example, the following survival

outcome that is like Survival except that the last generation has the same high average wellbeing as the previous generations.

Survival $\oplus$ (5 6 7 8 9 9 9 ... 9 9)

Since the worst-off generation in Survival $\oplus$ is better off than the worst-off generation in Alternative and its improved versions, we cannot apply Weak Impartiality. Nevertheless, some improved versions of Alternative are clearly better than Survival $\oplus$. We may give extra weight to the welfare of the worst off, but even welfare increases for the worst off can be outweighed by sufficiently large welfare increases for sufficiently many other people. Most plausibly, for example, Alternative $^+$ is better than Survival $\oplus$: the worst-off generation is better off in Survival $\oplus$ but *all* other generations are better off to a *much* larger extent in Alternative $^+$.

Since Alternative could always be further improved by increasing the wellbeing of all generations except the present one without becoming better than Extinction according to the neutrality view, there is just no version of Survival that the neutrality view could hold to be better than Extinction. The only way to stop this escalation is to restrict the neutrality view: if additions of happy people become good once their wellbeing surpasses a certain level, improving Alternative will eventually result in an outcome that is better than Extinction. But as long as there is neutrality, Alternative can be improved without becoming better than Extinction and while becoming better than additional versions of Survival.

4. No Badness despite the Loss of Non-Welfarist Values

I have argued that the harms to the present generation in Extinction do not allow the neutrality view to hold that Extinction is worse than Survival. This is in itself a surprising

and unpalatable result. However, it does not mean that the neutrality view cannot hold that Extinction is worse than Survival *all things considered*: Extinction could still be worse in virtue of non-welfarist values. In fact, it is a common view that the badness of human extinction is at least partially explained by the fact that non-welfarist values favour humanity's survival. In this section, I will argue that even if there are such non-welfarist values, the neutrality view needs to accept that Extinction is not worse than Survival. That is because it is still the case that if the neutrality view is true, Alternative is not better than Extinction but better than Survival *if* Alternative is interpreted in a certain way.

Before restating my argument, let us have a closer look at the non-welfarist values that have been held to favour humanity's survival over its near-term extinction. It will be useful to divide these values into two broad categories: those that favour human survival necessarily and those that only favour certain cases of human survival. According to values of the first category, the fact that humanity will continue existing for longer if it survives is necessarily good in some non-welfarist respect.¹⁰⁵ There are several intimately related non-welfarist values that have been proposed to explain this. Among them are the value of humanity's existence or of humanity itself, the value of species with certain attributes (such as rationality) that humanity exhibits, and the value of biodiversity (e.g. Hurka 1983; Frick 2017; Henning 2022, 143). It is not easy to disentangle these values and it often remains unclear which of them exactly is held to explain the badness of human extinction.¹⁰⁶ But the details won't matter for my argument. The general idea that unifies these values is that humanity's continued existence might be valuable in the same way in which the continued existence of (at least some) other animal species might be valuable, which is a rather

¹⁰⁵ Sometimes the qualification is added that humanity must continue existing in a *flourishing state* (e.g. Frick 2017, 360). I will include values that only favour humanity's survival under this condition in the first category.

¹⁰⁶ See Lenman (2002) who carefully distinguishes between these values. He adds the value of plenitude as another possible explanation.

common view. Additionally, the continued existence of humanity might be *especially* valuable because of humanity's special attributes such as its rationality or intelligence. Whatever ultimately explains it, I will refer to the value that is necessarily realized if humanity survives as "the value of humanity's presence".

In the second category are non-welfarist values that favour certain cases of human survival but not human survival *per se*. Generally speaking, these are values that will be realized in case of humanity's survival *if* future generations end up doing or achieving certain things. Even though it is not necessary that future generations will do these things, we may well expect this and therefore expect human extinction to be bad. It has been argued, for example, that human extinction could be bad because it would prevent all the great achievements in science, arts, and morality that we can expect from future generations (Parfit 1984, 454; Ord 2020, 305). Another popular view is that human extinction would be bad because it would prevent the completion of various unfinished projects that humanity started in the past, so that there would be "unfinished business" (Bennett 1978; Knutzen 2023).¹⁰⁷ In principle, many other (supposed) non-welfarist values could be added to this list. It has been held, for example, that "impersonal value [...] can be found in friendship, love, altruism, knowledge, perfection, beauty, rights, duty, virtue, and truth" (Temkin 2003, 779). As far as future generations would realize these values, they could contribute to explaining the badness of human extinction.

There is, then, a long list of potential non-welfarist values that might favour humanity's survival over its extinction.¹⁰⁸ Could the neutrality view accommodate some of

¹⁰⁷ This is different from the aforementioned view, which could be characterized as *perfectionist*. As Knutzen (2023, 13) explains, the unfinished-business view "does not imply that it would be good for there to be more meaningful achievements in the history of the universe". According to that view, achievements only have value insofar as they amount to the completion of an unfinished project that humanity has already started.

¹⁰⁸ It is of course not obvious that any of these actually *are* non-welfarist values. See, for example, Crisp (2003, 747) who rejects non-welfarist values in general. See also Fanciullo (2024) who argues that the badness of human extinction should not be explained in terms of non-welfarist values since there is no justified weighting of these values against the disvalue of bad lives. For the sake of argument, I will assume that these objections can be met.

these values? Given that the view only states that additional happy people do not make the world better *merely because* they would have good lives, the neutrality view is in principle compatible with all kinds of non-welfarist values. However, advocates of the neutrality view generally want to maintain that additions of people to an already existing generation – unlike additions of people who form *new* generations – are not good but neutral. Unlike *diachronic* additions, *synchronic* additions, as Lenman (2002) puts it, are not supposed to be valuable by default. As far as the neutrality view is to account for this judgement, it should not explain the badness of human extinction based on non-welfarist values that equally favour synchronic and diachronic additions. For example, the badness of human extinction should then not be explained by a view like Temkin's that assigns value to various products of common human conduct, such as friendship and love.

More generally, it seems that virtually *all* non-welfarist values that could explain the badness of human extinction could also be realized by, and therefore favour, synchronic additions. People don't have to extend humanity's presence to make great achievements or to advance humanity's uncompleted projects, for example. Even the value of humanity's presence itself could favour synchronic additions: a larger population could make humanity more resilient against extinction risks.¹⁰⁹ However, some of these non-welfarist values arguably tend to favour diachronic additions *more* than synchronic additions. For example, once the population is sufficiently large, as it currently seems to be, additional people don't significantly (if at all) increase the expected duration of humanity's existence. Similarly, it could be argued that future people are more likely to make great scientific achievements and that they will be in a better position to complete humanity's projects than additional

¹⁰⁹ Perhaps, the value of humanity's presence even favours synchronic additions *necessarily* because every plausible specification of that value does not discriminate between synchronic and diachronic additions. See Lenman (2002) who makes this argument and Frick (2017) who rejects it. I will assume, for the sake of argument, that this problem can be avoided by an appropriate specification of the value of humanity's presence.

people in the current generation. Explanations for the badness of human extinction in terms of these non-welfarist values therefore don't necessarily translate into explanations for the value of synchronic additions. For the sake of argument, I will assume that the neutrality view can indeed accommodate some non-welfarist values, including the value of humanity's presence itself, that favour Survival over Extinction.

In what follows, I will argue that the neutrality view cannot plausibly hold that Survival is better than Extinction, nevertheless. My strategy will be to extend my previous argument to the case of non-welfarist values as follows: I will first interpret Alternative in a way that allows me to stipulate that non-welfarist values do not favour Alternative over Extinction. Given this stipulation, the premise that Alternative is not better than Extinction if the neutrality view is true can be maintained. At the same time, this stipulation casts doubt on whether Alternative is better than Survival, which is the other substantive premise of my argument. I will therefore argue in a second step that this premise remains plausible even with the new stipulation in place. So my key claim will be that, even if there are non-welfarist values that favour Survival over Extinction that the neutrality view can accommodate, there is a version of Alternative that is both i) not better than Extinction according to the neutrality view (*because* this version of Alternative is not better in non-welfarist terms) and yet ii) better than Survival (*even though* it is not better in non-welfarist terms). Just as before, we can then infer by the transitivity of 'better than' that Survival is not better than Extinction if the neutrality view is correct. I will begin by specifying the relevant version of Alternative and then argue for the two parts of my key claim.

As a reminder, here are the welfare distributions of the three outcomes that my argument compares.

Extinction	(2	/	/	/	/	/	/	...	/	/)
Alternative	(2	10	10	10	10	10	10	...	10	10)

Survival (5 6 7 8 9 9 9 ... 9 2)

So far, I have interpreted Alternative as another survival outcome in which there are as many future generations as in Survival. This interpretation is not *implied* by the above representation of Alternative's welfare distribution since the numbers are ordered according to when the people whose welfare they represent exist in *Survival* (to ensure that numbers in one column represent the wellbeing of the same individuals). That these people also form consecutive future generations in Alternative is only one possible interpretation that is intuitive since Alternative contains exactly the same people as Survival. But since we are now considering the possibility of non-welfarist values that might favour survival outcomes over human extinction, I will need to reinterpret Alternative. Otherwise, Alternative might be overall better than Extinction in virtue of non-welfarist values that the neutrality view could accommodate. To maintain my premise that Alternative is not better than Extinction if the neutrality view is true, I will therefore interpret Alternative such that the additional people (i.e. everyone with wellbeing 10) do not form future generations as in Survival. They are the same people as in Survival, but they live along with other people in the past and thereby don't prolong the existence of humanity at all. One way to imagine this is that each generation in Alternative has many more descendants, many of whom have accordingly different parents than in Survival. If people could not possibly have different parents, we can imagine alternatively that the people in Alternative live in the same generational order as in Survival but in an overall shorter period of time, which is made possible by earlier conception in each generation.¹¹⁰ Either way, humanity goes extinct in

¹¹⁰ See Ord (2020, 194, Fn. 19) for some more concrete technological possibilities. If you think that it is metaphysically impossible that the same people who could form future generations could also form parts of past generations, you can imagine that the people in Alternative resemble the people in Survival as closely as possible (e.g. in terms of their appearances, beliefs, values, and goals) but are not identical to them. I think that this would not affect the plausibility of my subsequent arguments by much.

Alternative at the same time at which it goes extinct in Extinction. In other words, Alternative is only ‘synchronically more numerous’ than Extinction.

This is the type of case in which advocates of the neutrality view normally hold that additional happy people do not make the world better. As discussed above, even synchronic additions could be favoured by certain non-welfarist values, such as the value of scientific achievements. But I will stipulate that this is not the case. More precisely, I will stipulate that they do not realize non-welfarist values to the degree that the world becomes overall *better* with regard to these values. That does not necessarily mean that *no* additional instances of these values are realized by the additional people in Alternative. That depends on how the relevant non-welfarist values are aggregated. If, for example, what matters is the average of these values per person, the additional people in Alternative could bring about as many instances of these values as the other people. And if what matters are only the greatest instances of these values, as Nietzsche seems to have held with regard to achievements (see Hurka 2007, 18), the additional people in Alternative could bring about any instances of these values that are less great than the greatest instances of these values.¹¹¹ Either way, what matters is that Alternative is not better than Extinction in non-welfarist respects. Given that Alternative is only synchronically more numerous than Extinction, this should be at least a *possibility* according to the neutrality view.

Improved versions of Alternative, such as Alternative+, could also be interpreted synchronically. Analogously, they can also be stipulated to be no better than Extinction in non-welfarist respects. It could be objected that some very high level of welfare can only be reached by individuals who also realize some non-welfarist values, so that Alternative cannot be indefinitely improved in welfarist terms without also making it better in some

¹¹¹ Even if what matters is the *total* of non-welfarist values, we could imagine that the additional people bring about some instances of these values. This can result in the same total if the other people bring about less instances than in Extinction: the same total could be stretched out over more people.

non-welfarist respects. On a view like Temkin's that assigns non-welfarist value to various products of human conduct, this has some plausibility. Perhaps, for example, lives cannot be arbitrarily good without containing love and friendship, which Temkin takes to have non-welfarist value. However, I have already argued that the neutrality view should not explain the badness of human extinction based on such an abundance of non-welfarist values to salvage some degree of neutrality in the case of synchronic additions. Moreover, even on a view like Temkin's it is plausible that very high degrees of welfare, if not *any* degree of welfare, can be reached without realizing any non-welfarist values. There seem to be at least some welfarist values, such as pleasure, desire satisfaction and achievements that are only personally important, whose realization does not require anything with non-welfarist value. It seems possible to reach very high welfare levels, if not any welfare level, by scaling up these welfarist values while removing all suffering.

I conclude that, if interpreted synchronically, Alternative and its improved versions could be no better than Extinction in non-welfarist respects if the neutrality view is true. I will stipulate that this is indeed the case: Alternative and its improved versions are not better than Extinction in non-welfarist respects. Since Alternative and its improved versions are, according to the neutrality view, not better than Extinction in welfarist respects *either*, Alternative and its improved versions are not overall better than Extinction if the neutrality view is true. This reinstates the first substantive premise of my argument.

At the same time, my stipulation that Alternative is not better than Extinction in non-welfarist respects casts doubt on my other substantive premise, according to which Alternative is overall better than Survival. After all, I have granted that Survival *is* better than Extinction in non-welfarist respects. This makes it plausible that Survival is also better than Alternative in non-welfarist respects. Given my synchronic interpretation of Alternative, this holds at least for the value of humanity's presence: since humanity goes extinct in Alternative at the same time as in Extinction, humanity will be present for much

longer in Survival than in Alternative. Other non-welfarist values, such as the value of achievements or the completion of humanity's projects, might additionally favour Survival over Extinction and Alternative, depending on how Survival is specified. In what follows, I will argue that *even if* Survival is additionally favoured by some of these values, Alternative (or some improved version of it) is better than Survival.¹¹²

Compare again the welfare distributions of the two outcomes.

Alternative	(2	10	10	10	10	10	10	...	10	10)
Survival	(5	6	7	8	9	9	9	...	9	2)

I have argued before that Alternative is at least as good as Survival as far as the welfare of the present and the last generation is concerned: the welfare gains for the last generation in Alternative outweigh, or at least compensate, the welfare gains for the present generation in Survival. This basically remains true on the synchronic interpretation of Alternative. The only difference is that, on this interpretation, the groups of people whose welfare we compare do not constitute future generations in Alternative. But we are still comparing the same welfare differences for the same groups of people. So we only need to specify that Alternative is at least as good as Survival as far as the welfare of the people is concerned who form the present and the last generation *in Survival*. What remains is that all other people have better lives in Alternative and that Survival is favoured by some non-welfarist values. What is more important?

Consider again the welfare gains in Alternative. They are for many billions (if not *trillions*) of people – almost all the people who form future generations in Survival are much better off in Alternative. To give an idea of the welfare difference between 10 and 9, I said earlier that this difference is about as large as the welfare difference that results from

¹¹² I don't *need* to assume this since my assumption (7) only requires that Survival is favoured by non-welfarist values at all.

several additional years of life with a high quality; the difference between 10 and 8, 7, and 6 is accordingly larger yet. At the same time, these large differences in wellbeing do not *have* to be the result of longer lives in Alternative; they could also be the result of less suffering. This could be, for example, because diseases and natural catastrophes are much less frequent in Alternative. Intuitively, this reduced suffering for many billions of people outweighs even impressive non-welfarist values. At the least, it seems to outweigh the value that is grounded in the extended presence of the human species in Survival: it would be better if the future generations in Survival lived along with past generations if that meant that their lives would contain much less suffering. This seems to remain true even if we assume that the future generations in Survival create additional non-welfarist value by, for example, making some great scientific discoveries and advancing some uncompleted cultural developments. There may be some point at which these instantiations of non-welfarist values become so numerous, or individually so impressive, that they cancel out the welfare gains in Alternative. But many combinations of important non-welfarist values that go beyond what Survival contains necessarily seem clearly outweighed by the welfare gains in Alternative.

This argument can be reinforced by increasing the welfarist stakes. One way of doing this is to employ an improved version of Alternative, such as Alternative+, in which the people who form the future generations in Survival are better off yet. Since the welfare differences between Alternative+ and Survival are much larger than the welfare differences between Alternative and Survival, even greater and more numerous instantiations of non-welfarist values in Survival can be outweighed by the welfare gains in Alternative+. Perhaps, for example, Alternative is not better than Survival if the future generations in Survival produce humanity's greatest works of art and build humanity's most just societies. But even these non-welfarist values are plausibly outweighed by the immense welfare gains in Alternative+ or a yet more improved version of Alternative.

Another way of increasing the welfarist stakes is to increase the number of people that form the future generations in Survival and the corresponding groups of people in Alternative.¹¹³ We could imagine, for example, that the wellbeing levels assigned to Survival and Alternative do not each represent the wellbeing of about 10 billion people, but the wellbeing of 10 *trillion* people. In these scaled-up versions of Alternative and Survival, there are many more people who are better off in Alternative than in Survival. The individual welfare differences remain equally large, but they affect now far more people: for each individual who was better off in Alternative, there are now 1000 people who are better off in Alternative. Since the welfare gains in the scaled-up Alternative are more numerous, they can outweigh additional non-welfarist value.

It seems that there is no limit to increasing the welfarist stakes in these ways. In particular, the difference in welfarist value between Alternative and Survival could always be further increased by scaling-up the two outcomes. Even if there are already trillions of people who are better off in Alternative than in Survival, the value difference between Alternative and Survival increases if there is yet another individual who has a better life in Alternative than in Survival. Arguably, the same holds for increasing the welfare levels in Alternative: the value difference between Alternative and Survival could always be increased by further improving the lives of the people in Alternative. It seems plausible, therefore, that *any* amount of non-welfarist value in Survival could be outweighed by *some* amount of welfarist value in Alternative. That would not mean that welfarist values are *superior* to non-welfarist values. For one, it could be the case that the extent to which non-welfarist value could favour Survival is bounded. At least with regard to some non-welfarist values, this is plausible. As Knutzen (2023, 13) argues, for example, the value of

¹¹³ Scaling up Survival and Alternative in this way (rather than by adding new generations) does not imply that the *non*-welfarist stakes increase as well since Survival and Alternative only become *synchronically* more numerous.

completing unfinished projects has an upper bound given the actual history of humanity: once all projects are completed, there will be no unfinished business left that could be finished. But it is not necessary for my claim that non-welfarist values, or their potential support for Survival, are bounded. Both welfarist and non-welfarist values could be unbounded, so that any amount of value can be reached by sufficiently many of their instantiations. In that case, there would be no *single* amount of welfarist value that can outweigh each amount of non-welfarist value. But it could still be the case that each amount of non-welfarist value can be outweighed by *some* amount of non-welfarist value. It would then just be symmetrically the case that also any amount of *welfarist* value can be outweighed by some amount of *non-welfarist* value.

If this argument is sound, some version of Alternative is better than some version of Survival *no matter* the extent to which non-welfarist values favour Survival. Accordingly, since Alternative and its discussed versions are not better than Extinction if the neutrality view is true, the neutrality view could then not hold that Survival is better than Extinction *whatever* we assume about the non-welfarist values in Survival. But even if this argument fails, the neutrality view needs to accept that Survival is not better than Extinction even if Survival is *strongly* favoured by non-welfarist values.

5. Conclusion

The convergence view is false. If the neutrality view is true, near-term human extinction would not be bad even under very optimistic assumptions about humanity's future. Even the harms to the last generation and the loss of non-welfarist values do not allow the neutrality view to hold that human extinction is overall bad. As mentioned before, this doesn't mean that there is no moral reason to prevent human extinction at all. But it does mean that, if the neutrality view holds, there is no reason to prevent human extinction that

is grounded in its badness. This plausibly makes a big difference to the *overall strength* of our reason to prevent human extinction: if our reason to prevent bad outcomes roughly scales with the outcome's badness, we'd have a very strong moral reason to prevent extinction if extinction was terribly bad. Now we learn that, against the convergence view, human extinction wouldn't be bad at all if the neutrality view holds. So the neutrality view means that we lose a potentially very strong reason to prevent human extinction.

But this is not the only upshot of the argument. We also learn that the neutrality view has absurd implications. After all, human extinction seems clearly bad in the particular cases that I discussed. That the neutrality view denies the badness of human extinction in these cases is an excellent reason to reject the view.¹¹⁴

Could the neutrality view be amended to avoid this reductio? One possibility is to hold that it is only *normatively* neutral to make happy people: outcomes with additional happy people are (sometimes, at least) better, but we have no reason to bring about these outcomes. As many advocates of the neutrality view have noted, however, this combination of views is hard to justify (e.g. Bader 2022, 16; Frick 2022, 234). It requires that there is not even a *pro tanto* reason to choose options with better outcomes. That is implausible (*pace* Sinclair 2018). If there is no axiological neutrality, there won't be normative neutrality either.¹¹⁵

Another possibility is to restrict the neutrality view so that additions of happy people become good once they surpass a certain wellbeing level; they are only neutral if their wellbeing falls into a bounded "neutral range" (Broome 2022, 399). According to this restricted neutrality view, versions of Alternative in which the wellbeing of the additional

¹¹⁴ So why should we care about the just mentioned normative implications of the neutrality view? Because we shouldn't be *certain* that the neutrality view is false, and because we shouldn't ignore moral views that deserve some credence (see MacAskill, Bykvist, and Ord 2020).

¹¹⁵ At least, there will be a reason to create happy people with *justifying* strength; maybe this reason doesn't moreover have any *requiring* strength (Thomas 2023, 491; see also Pummer 2024).

people surpasses the neutral range are better than Extinction and therefore could not be used for my argument. This means that there will be *some* possible survival outcomes that are better than extinction. But in many other cases, the restriction to a neutral range won't make a difference, provided that it includes at least some high wellbeing levels. As I argued, the unimproved Alternative is better than Survival, so no improved version of Alternative is needed to argue that Survival is not better than Extinction. Moreover, even *degraded* versions of Alternative, in which the additional people have a wellbeing below 10, could be used to argue that degraded versions of Survival are not better than Extinction. As long as the degraded versions of Survival still meet the conditions (1) to (7), which only require a positive average wellbeing that is higher than in the last few generations, this suffices for reductio.¹¹⁶

Another popular amendment of the neutrality view is to hold that additions of happy people can be *bad* (even though never good) if they are worse off than in an available alternative (e.g. Frick 2022; Roberts 2022). Due to this dependence on available alternatives (or “options”), it has been argued that transitivity only holds within option-sets (ibid.). To retain a valid argument, I would therefore need to stipulate that the same option-set is available in each comparison that the argument makes. Here is one obvious way to do this: in each comparison, the option-set contains Extinction, Alternative, and Survival. We can thereby make the valid inference that, according to the (amended) neutrality view, Extinction is not worse than Survival when Alternative is an available option. This qualification barely mitigates the reductio: the additional people in Survival are even better off in Alternative, but claiming that Survival is therefore not better than Extinction is absurd. Even if the availability of Alternative negatively affects the value of Survival, this

¹¹⁶ Rabinowicz' (2022) view on which the neutral range comprises only lives that are neither better nor worse than non-existence, and hence comprises *no* positive wellbeing levels, is perhaps the only view that can robustly avoid this type of reductio. Of course, at this point there is not much – or rather: nothing – left of being ‘neutral about making happy people’.

negative effect is clearly outweighed by the values that favour Survival over Extinction, such as the immense welfare gains for all present people.

There is one more possibility: agreeing that my arguments show what the neutrality view would need to accept in order to hold that Extinction is worse than Survival, and to accept some of these costs. In particular, one could endorse a *Nietzschean* neutrality view, according to which certain non-welfarist values are strongly superior to welfarist values. To robustly avoid the verdict that Extinction is not worse than Survival, however, this superiority would need to be of the most extreme kind. Merely claiming that there are *some* constellations of non-welfarist values that can outweigh all welfarist considerations would only enable holding that Extinction is worse than Survival *if* Survival contains these amazing constellations of non-welfarist values. In all other cases, Extinction would still be no worse than Survival. To actually avoid the reductio, Nietzschean neutrality would therefore need to hold that even the *smallest* possible extent to which non-welfarist values could favour Survival over Extinction is enough to outweigh all welfarist considerations. And besides Nietzsche himself, perhaps, nobody seems to have such a strong stomach.

References

- Bader, Ralf M. 2022. "The asymmetry." In *Ethics and Existence: The Legacy of Derek Parfit*, edited by Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan. Oxford: Oxford University Press.
- Beckstead, Nicholas. 2013. "On the Overwhelming Importance of Shaping the Far Future." PhD dissertation, Rutgers University.
- Bennett, Jonathan. 1978. "On Maximizing Happiness." In *Obligations to Future Generations*, edited by Richard Sikora and Brian Barry, 61–73. Philadelphia: Temple University Press.
- Bostrom, Nick. 2013. "Existential Risk Prevention as Global Priority." *Global Policy* 4 (1): 15–31.
- Broome, John. 2004. *Weighing Lives*. Oxford: Oxford University Press.

- . 2005. "Should We Value Population?" *Journal of Political Philosophy* 13 (4): 399–413.
- . 2022. "Climate Change and Population Ethics." In *The Oxford Handbook of Population Ethics*, edited by Gustaf Arrhenius et al. <https://doi.org/10.1093/oxfordhb/9780190907686.001.0001>.
- Crisp, Roger. 2003. "Equality, Priority, and Compassion." *Ethics* 113 (4): 745–763.
- Fanciullo, James. 2024. "Why prevent human extinction?" *Philosophy and Phenomenological Research* 109 (2): 650–662.
- Finneron-Burns, Elizabeth. 2017. "What's wrong with human extinction?" *Canadian Journal of Philosophy* 47 (2-3): 327–343.
- Francis, Tomi. 2021. "Anonymity and Non-Identity Cases." *Analysis* 81 (4): 632–639. doi: 10.1093/analys/anab031.
- . 2022. "Totalism." DPhil Thesis, University of Oxford.
- Frick, Johann. 2017. "On the survival of humanity." *Canadian Journal of Philosophy* 47 (2-3): 344–367. <http://dx.doi.org/10.1080/00455091.2017.1301764>.
- . 2022. "Context-Dependent Betterness and the Mere Addition Paradox." In *Ethics and Existence: The Legacy of Derek Parfit*, edited by Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan. Oxford: Oxford University Press.
- Greaves, Hilary, and William MacAskill. 2025. "The Case for Strong Longtermism." In *Essays on Longtermism*, edited by James Barrett, Hilary Greaves, and David Thorstad. Oxford: Oxford University Press.
- Henning, Tim. 2022. *Die Zukunft der Menschheit – soll es uns weiter geben?* Berlin: J.B. Metzler.
- Hurka, Thomas. 1983. "Value and Population Size." *Ethics* 93 (3): 496–507.
- . 2007. "Nietzsche: Perfectionist." In *Nietzsche and Morality*, edited by Brian Leiter and Neil Sinhababu, 9–31. Oxford: Oxford University Press.
- Kaczmarek, Patrick, and Simon Beard. 2020. "Human extinction and our obligations to the past." *Utilitas* 32 (2): 199–208.
- Knutzen, Jonathan. 2023. "Unfinished Business." *Philosophers' Imprint* 23 (4). doi: 10.3998/phimp.1441.
- Lenman, James. 2002. "On Becoming Extinct." *Pacific Philosophical Quarterly* 83: 253–269.
- MacAskill, William, Krister Bykvist, and Toby Ord. 2020. *Moral uncertainty*. Oxford: Oxford University Press.

- . 2022. *What We Owe the Future*. Oneworld Publications.
- Mogensen, Andreas. 2025. "Would a World Without Us Be Worse?" In *Essays on Longtermism*, edited by James Barrett, Hilary Greaves, and David Thorstad. Oxford: Oxford University Press.
- Narveson, Jan. 1973. "Moral problems of population." *The Monist* 57 (1): 62–86.
- Ord, Toby. 2020. *The precipice: existential risk and the future of humanity*. New York: Hachette Books.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Pummer, Theron. 2024. "Future Suffering and the Non-Identity Problem." *GPI Working Paper* No. 17-2024. <https://globalprioritiesinstitute.org/future-suffering-and-the-nonidentity-problem-theron-pummer/>.
- Rabinowicz, Wlodek. 2009. "Broome and The Intuition of Neutrality." *Philosophical Issues* 19: 389–411.
- . 2022. "Getting Personal: The Intuition of Neutrality Reinterpreted." In *The Oxford Handbook of Population Ethics*, edited by Gustaf Arrhenius et al. Oxford Handbooks. <https://doi.org/10.1093/oxfordhb/9780190907686.013.17>.
- Roberts, Melinda. 2022. "The Person-Based Intuition and the Better Chance Puzzle." In *The Oxford Handbook of Population Ethics*, edited by Gustaf Arrhenius et al. Oxford Handbooks. <https://doi.org/10.1093/oxfordhb/9780190907686.001.0001>.
- Scheffler, Samuel. 2018. *Why Worry About Future Generations?* Oxford: Oxford University Press.
- Sinclair, Thomas. 2018. "Are we conditionally obligated to be effective altruists?" *Philosophy & Public Affairs* 46 (1): 36–59.
- Steele, Katie. 2025. "Longtermism and Neutrality about More Lives." In *Essays on Longtermism*, edited by James Barrett, Hilary Greaves, and David Thorstad. Oxford: Oxford University Press.
- Temkin, Larry S. 2003. "Egalitarianism defended." *Ethics* 113 (4): 764–782.
- . 2008. "Is living longer living better?" *Journal of Applied Philosophy* 25 (3): 193–210.
- Thomas, Teruji. 2023. "The asymmetry, uncertainty, and the long term." *Philosophy and Phenomenological Research* 107 (2): 470–500.

Chapter 4: The Moral Importance of Low-Welfare Species

Abstract

Many species seem to have much smaller welfare ranges than we do. Which importance should we assign to the welfare of these *low-welfare species* when we have to decide whether to benefit members of these species or other humans? In particular, should we ever prioritize low-welfare species when some of the most significant human goods are at stake? It could be argued that benefits for low-welfare species are *irrelevant* to significant human benefits because they are much smaller in comparison. I begin by developing this argument, but subsequently reject it based on the consideration of a hypothetical *high-welfare species* whose welfare range vastly exceeds the human welfare range. Arguments from empathy and fairness are given to support the opposite view on which some benefits for low-welfare species can outweigh even significant human benefits. Finally, I propose a principle that reconciles this view with the intuition that human headaches cannot outweigh human lives.

1. Introduction

Many species seem to have welfare ranges that are much smaller than the welfare range of our species: the difference between the maximum and minimum level of lifetime wellbeing that members of these species can achieve is much smaller than in our case. Given that they have any welfare at all, this seems particularly plausible for invertebrates; but also many (small) mammals, fish, reptiles and birds plausibly have much smaller welfare ranges than we do.¹¹⁷ There are various reasons this might be the case: members of these species have

¹¹⁷ See Rethink Priorities' pioneering research on the welfare ranges of other species, in particular Fischer (2023, 2024). Note that on their definition, welfare ranges measure individuals' *momentary* welfare rather than their lifetime welfare.

much shorter lives, they have less developed brains that might only allow for less intense pleasures and pains, and they might not be able to attain certain sophisticated goods (such as scientific understanding) that require a high level of intelligence or rationality. Generally speaking, the physical and psychological constitution of many species seems to limit their capacity for wellbeing in a way that does not allow them to be nearly as well off or badly off as we can be, at least under any realistic circumstances.¹¹⁸ I will refer to these species whose welfare ranges are much smaller than (or “tiny” compared to) the human welfare range as “low-welfare species”.¹¹⁹

In this essay, I will discuss the moral importance of low-welfare species. More specifically, I will ask whether we should always prioritize some significant human benefits over benefits for members of low-welfare species, or whether some benefits for members of low-welfare species can, if numerous enough, outweigh even the most significant human benefits. For concreteness, I will often use the saving of a human life as an example for one of the most significant human benefits, where the saved life is assumed to be worth living and of substantial length. More specifically, we can assume that the life would be about as good and long as a human life can be, and that the person living it has only had a short life so far. My question can then be stated as follows: Is there any number N and benefit B such

¹¹⁸ The addition of “under any realistic circumstances” is supposed to allow for the possibility that members of low-welfare species could be much better off in some circumstances that are fundamentally different from ours. Perhaps, for example, mice could live for many hundreds of years in extreme pleasure if there was a life-prolonging experience machine that is optimized for mice (cf. Nozick 1974, 42–45). While this might be a metaphysical (and perhaps nomological) possibility, it is not a realistic one. Hence, this possibility does not affect the welfare range of mice. That is not to say that it is always easy to specify whether a possible circumstance should be considered “realistic” or “fundamentally different”. For example, should the (mere) possibility of radical enhancement of one’s welfare-relevant capacities count as a realistic circumstance? My tentative answer to this question is that such circumstances should be considered “fundamentally different” given that they are not actually an available technology. See Vallentyne (2005) for an insightful discussion.

¹¹⁹ To be a bit more precise, I am thinking of welfare ranges that are smaller than the human welfare range *by at least several orders of magnitude*. A species with a welfare range that is about half as large as the human welfare range would not qualify as a low-welfare species; a species with a welfare range that amounts to less than 0.1% of the human welfare range would. I am not assuming that such comparisons could be made precisely, even in principle. But I do assume that welfare ranges are at least imprecisely comparable, so that we can say, for some N , that one welfare range is *at least* N times as large as another welfare range even if we cannot say, for *any* N , that it is *exactly* N times as large.

that one ought to confer benefit B to N members of a low-welfare species rather than save one human life?

While it seems obvious that we should save a human life rather than benefit a single member of a low-welfare species (whatever the latter benefit may be), it is unobvious whether the same is true for *any* number of any benefits for members of low-welfare species. At the same time, this question is of high practical importance. As noted above, many actual species seem to be low-welfare species and, moreover, many of these species – in particular invertebrate species – contain incredibly many individuals. Of course, we cannot choose to benefit an arbitrarily large number of these “small animals” that are plausible candidates for members of low welfare species. But it seems that we can benefit *far* more of them than we can benefit humans. That goes in particular for the enormous numbers of small animals in factory farms (such as fish, shrimp, and farmed insects) whose living conditions we could greatly improve. But also the suffering of an even much larger number of small animals in the wild provides an opportunity for us to potentially do a lot of good. Even if the tractability of benefitting small wild animals is quite low, their astronomical numbers could render interventions that are aimed at their wellbeing the most effective (in terms of the increase in total wellbeing per resource).¹²⁰ If benefits for low-welfare species could in principle outweigh all human benefits, it may well be, therefore, that they sometimes actually do outweigh some of the most significant human benefits.

This conclusion would be hard to accept. It would mean that in decision situations in which we can save either humans or members of low-welfare species but (due to scarce resources) not both of them, we sometimes ought to save the members of low-welfare species while allowing the humans to be severely harmed. Most of us are likely to face

¹²⁰ See Horta & Teran (2023) for different promising ways of benefitting wild animals effectively. See Sebo (2023) for the argument that the sheer number of small animals makes it plausible that we sometimes ought to prioritize them on a utilitarian calculus.

decisions of this kind. For example, when deciding where to donate some of our money, we could use that money to protect people in extreme poverty from deadly tropical diseases, or we could use the money to fight the inhumane treatment of factory-farmed salmon. Of course, we could split our donations between the two causes or look for a third option that allows us to benefit both humans and small animals. But usually, for every dollar we donate, there will be some trade-off between human and small animal welfare; there is no perfect correlation between what is best for humans and what is best for small animals. What goes for money also goes for other resources. For example, we could use some of our spare time to protest against the discrimination of certain human minorities, or we could use that time to advocate for improving the living conditions of bees. If benefits for low-welfare species can in principle outweigh all benefits for humans, it might well be (although this is far from certain!) that we should advocate for the bees rather than for the discriminated people and that we should donate to help salmon rather than people in extreme poverty.¹²¹

As I said, these conclusions would be hard to accept; they may also seem unintuitive. But we should be wary of rejecting them without substantive argument. Not only could our intuitive rejection of these practical conclusions merely be the result of motivated reasoning that is based on our self-interest or speciesism; their unintuitiveness could also be explained by the underlying unintuitive empirical premises about the effectiveness of benefitting small animals such as the (factually correct) premise that there are so incredibly many small animals. Even if we accept these empirical premises, they are hard to grasp intuitively and may therefore not be properly reflected in our judgement. We

¹²¹ One important complication (besides many others) for reaching these conclusions is the non-identity problem (Parfit 1984, ch.17): when trying to benefit small animals on a large scale by improving their living conditions, we will probably also change their identities. We would therefore replace less well-off animals by better-off animals rather than make particular animals better off. This could be held to make a moral difference. But that moral difference should not be overstated: we also change the identities of humans by many interventions such as those that mitigate climate change.

should therefore inquire whether there are good arguments for the view that some significant benefits for humans always outweigh benefits for low-welfare species.

I will present such an argument in the following section. Based on the consideration of species with welfare ranges that are much *larger* than our welfare range, I will, however, reject this argument (section 3). Subsequently, I will give two arguments for the opposite view on which some benefits for low-welfare species can outweigh even the most significant human benefits. The first of these arguments is based on the value of fairness (section 4), and the second is an argument from empathy (section 5). Finally, I discuss principles that could explain why benefits for low-welfare species, but not small benefits for humans, can outweigh human lives (section 6). I conclude in section 7.

2. The Argument from Irrelevance of Tiny Benefits

Here is a straightforward argument for the view that some benefits for humans cannot be outweighed by any benefits for low-welfare species: The welfare range of an individual's species restricts how much individuals of that species can be benefited: increasing their wellbeing from the minimum to the maximum wellbeing level of that species is the most that could be done for them. The welfare range of a low-welfare species is, by definition, tiny compared to the human welfare range. Hence, the largest benefits to individuals of a low-welfare species are tiny compared to the largest human benefits. Benefits to one individual cannot be outweighed by any number of comparatively tiny benefits to other individuals. Thus, no number of even the largest benefits to members of low-welfare species can outweigh any of the largest human benefits, such as the saving of a human life.

The crucial premise in this argument is the claim that benefits, no matter their number, cannot outweigh any benefit for another individual compared to which they are tiny. This claim, which I will refer to as "Irrelevance of Tiny Benefits", gains plausibility

from cases like *life for headaches* (see Norcross 1997): in a choice between saving a human life (which, as before, is assumed to be well worth living and of substantial length) and alleviating a huge number of mild headaches, it seems that we ought to save the human life, *no matter* the number of headaches.¹²² One way of expressing this intuition is to say that the alleviation of mild headaches is *irrelevant* when human lives are at stake (Scanlon 1998, 239). Irrelevance of Tiny Benefits provides a natural explanation for this intuition: the benefit of an alleviated headache is irrelevant to the benefit of a saved human life because the former benefit is tiny compared to the latter benefit. More generally, the comparative magnitude of benefits seems to explain our intuitions about whether one benefit is relevant to another benefit in many common cases. Intuitively, for example, cases of constant paralysis *are* relevant to the saving of a human life: for some number N , we should prevent N cases of constant paralysis rather than save a single human life (ibid.). The reason for this seems to be that the benefit of preventing constant paralysis, unlike the benefit of preventing a headache, is sufficiently large compared to the benefit of saving a life. Irrelevance of Tiny Benefits can be considered an instance of the more general view that the comparative magnitude of benefits determines their relevance.¹²³

A straightforward way of specifying this more general view is *Compare Absolute Magnitudes*: for all benefits B_1 and B_2 , B_1 is relevant to B_2 if and only if the magnitude of B_1 is sufficiently large compared to the magnitude of B_2 . Given that comparatively *tiny*

¹²² Norcross himself, who introduced this case, ends up rejecting the intuition that we should prioritize the life over any number of headaches. Many others (e.g. Dorsey 2009) are convinced that our intuition about this case is correct.

¹²³ Since this view implies that the numbers of benefits sometimes matter and sometimes don't matter, it is *partially aggregative*. Alternatively, Irrelevance of Tiny Benefits could be motivated by the *non-aggregative* view that the numbers of benefits *generally* don't matter (e.g. Taurek 1977). Given this view, it would only need to be argued that no *single* instance of a benefit can outweigh any benefit compared to which it is tiny; it would follow that the same holds for *any* number of comparatively tiny benefits. The non-aggregative view, however, has highly implausible implications (see Parfit 1978). For instance, it would follow that if we should save a life rather than prevent a case of constant paralysis, we should also save a life rather than prevent one *billion* cases of constant paralysis. But intuitively, the numbers do make a difference here.

benefits are not ‘sufficiently large’, Compare Absolute Magnitudes implies Irrelevance of Tiny Benefits. The same is true for many variations of Compare Absolute Magnitudes that also reflect other factors besides comparative benefit magnitudes. For example, Compare Absolute Magnitudes could be supplemented by a further necessary condition for a benefit’s relevance according to which the benefit may not be based on immoral preferences (such as in the case of sadistic pleasures). Since comparatively tiny benefits do not meet the first necessary condition of having a sufficiently large comparative magnitude, they would still count as irrelevant on this variation of Compare Absolute Magnitudes. Another popular variation gives extra weight to benefits for individuals who are less well off than the other potential beneficiaries (e.g. Voorhoeve 2014). Benefits that are too small to count as relevant on the simple Compare Absolute Magnitudes can be relevant on this variation of the view if the beneficiaries have comparatively low welfare. However, that does not have to hold for *tiny* benefits. Arguably, tiny benefits should be considered irrelevant even if they are conferred to individuals with much lower welfare. Otherwise, mild headaches could outweigh human lives after all if the headache patients are much less well off than the person whose life is at stake.

Overall, then, Irrelevance of Tiny Benefits is supported by case intuitions and is implied by several general views about the relevance of benefits that seem plausible. Nevertheless, I believe that Irrelevance of Tiny Benefits should be rejected. One reason to doubt this premise is, of course, the possibility that there are no irrelevant benefits at all: perhaps *all* benefits can, if numerous enough, outweigh any other benefit. This is a possibility that should be taken seriously.¹²⁴ My rejection of Irrelevance of Tiny Benefits, however, does not rely on this contentious view. In what follows, I will argue that

¹²⁴ See, for example, Norcross (1997) and Horton (2018) for arguments for this *fully aggregative* view.

Irrelevance of Tiny Benefits should be rejected even on the assumption that there are irrelevant benefits.

3. The Problem of High-Welfare Species

Consider a hypothetical “high-welfare species” with a welfare range that is so large that the human welfare range is tiny in comparison. Perhaps individuals of that species live much longer or are able to experience much more intensely than humans, for example. Imagine that we could save the life of one individual of that species that is about as good as lives of that species can be; if we save its life, the individual’s lifetime wellbeing will be close to the maximum wellbeing level of that species while its lifetime wellbeing would be close to neutral if we don’t save it. This seems to be a morally weighty consideration, but it also seems that the greatest goods for humans – such as our health, freedom, and flourishing – are still important moral factors that can be morally decisive. According to Irrelevance of Tiny Benefits, however, there is no number of any benefits for humans such that one ought to prioritize them since all human benefits are tiny compared to the benefit of saving the individual of the high-welfare species. This seems to be the wrong result. Surely, for example, there is *some* number of human lives such that we should save them rather than the one individual of the high-welfare species.

To see how plausible and weak this claim is, compare it to the intuition that we ought not to give all of our resources to a “utility monster” that gains much more wellbeing from each resource than we do in total (Nozick 1974, 41). This widespread intuition that is frequently employed against utilitarianism is sufficient to argue against Irrelevance of Tiny Benefits: if we should sometimes benefit humans even though we could alternatively benefit the utility monster by a much larger magnitude, much smaller (or “comparatively tiny”) benefits can sometimes outweigh much larger benefits. While I take this to provide

some reason to reject Irrelevance of Tiny Benefits, I don't take it to be a decisive reason since the intuition that we sometimes ought to prioritize humans over the utility monster can be doubted. After all, the benefits for the utility monster are not only much larger than each individual human benefit but also larger than the *sum* of all human benefits. (Otherwise, utilitarianism would agree that we should not give all of our resources to the utility monster.) It could be held, therefore, that while human benefits could *in principle* outweigh the benefit to the utility monster, they do not actually outweigh it since they don't add up to a sufficiently large total benefit.

This answer to the utility monster case is compatible with my claim that benefits for members of high-welfare species can be outweighed by some human benefits that are comparatively tiny. For my claim, it is sufficient that some human benefits could outweigh the benefit to the utility monster *in principle*. That is, I only need to claim that we should prioritize the human beneficiaries over the utility monster if the human beneficiaries were much more numerous (while the benefit to the utility monster remained unchanged). Given that each human beneficiary is benefited in a significant way, *this* seems close to indubitable. In particular, it seems clear that we should prioritize the human beneficiaries if they were so many that the sum of their benefits is *larger* than the benefit to the utility monster. Assume, for example, that the benefit to the utility monster is about a *million* times larger than the largest human benefit and that, with a given amount of resources, we could benefit either the utility monster or one *billion* humans in these ways. Benefitting humans would then be about 1000 times *more* effective (in terms of the total benefit per resource) than benefitting the utility monster. Of course, the "utility monster" would then cease to be a real utility monster since it would not derive more wellbeing from each resource than we do in total. But it would still be an example of an individual of a high-welfare species to which Irrelevance of Tiny Benefits accords priority: due to its much larger welfare range, we can benefit it much more than we could possibly benefit an individual human being.

This verdict of Irrelevance of Tiny Benefits seems clearly wrong: even if we should give all of our resources to Nozick's utility monster, we should not spend all of our resources on such high-welfare species that we can benefit much less effectively than humans.

As I said, this claim is very weak. It is compatible with the view that we should prioritize one individual of a high-welfare species over a single human or even one million humans. It is also compatible with the view that, unit for unit, the wellbeing of high-welfare species matters more (e.g. due to their more developed psychological capacities) than our wellbeing (see Kagan 2019, 149). I only claim that we should not give *absolute* priority to high-welfare species whenever significant goods of theirs are at stake. It also seems that we don't need to be very suspicious about this claim: not only does it not imply priority for humans but merely lack of absolute priority for other beings; we also have barely any self-interest in that claim being true since, as far as we know, there are no species with welfare ranges that are much larger than our welfare range.¹²⁵ This is in contrast to the question at issue whether we sometimes have absolute priority over low-welfare species: those probably exist, and it is in our self-interest to have absolute priority over them.

I conclude that Irrelevance of Tiny Benefits should be rejected and that, accordingly, the above argument for the view that no benefits for low-welfare species can outweigh a human life is unsound. Moreover, the consideration of high-welfare species poses a general challenge for this view: to motivate it, we would need a principle according to which no benefits for low-welfare species can outweigh human lives while some benefits for humans *can* outweigh lives of high-welfare species. As argued above, Irrelevance of Tiny Benefits cannot meet this challenge. More generally, it is hard to see how the challenge could be met by *any* well-motivated (and similarly parsimonious) principle. An obvious

¹²⁵ Interestingly, there could be high-welfare species in the future due to the possibility of sentient artificial intelligences (Shulman & Bostrom 2021). But this is a very unusual view that most people don't believe in. So this possibility plausibly does not affect our intuitions.

response is that benefits for low-welfare species are much smaller than benefits for humans: while benefits for low-welfare species are small compared to human benefits and human benefits are small compared to benefits for high-welfare species, there is still the fact that some human benefits are large while all benefits for low-welfare species are small. In other words, some human benefits, unlike any benefits for low-welfare species, are above a certain threshold above which all benefits are (intuitively) large and therefore relevant.

While this response gives us the desired results, it is also suspiciously convenient and anthropocentric. Our judgement that all benefits for low-welfare species are small seems to merely reflect that they are small compared to the range of human benefits. We may ask ourselves where members of high-welfare species would set the threshold above which benefits count as large enough to be relevant. Presumably, they would set the threshold at the level of benefits that are relatively large for *them* and hence higher than the largest human benefits, which *they* would consider small. Even if there is a threshold above which all benefits are relevant, we have no good reason to believe that our placing of the threshold is the right one. We should therefore not rely on this proposal to explain why some human benefits are relevant to much larger benefits for high-welfare species.¹²⁶

In what follows, I will argue that, even when setting this challenge aside, there are good reasons to accept the opposite view on which some benefits for low-welfare species *can* outweigh human lives. I will present two arguments that attempt to make this view directly plausible. These arguments will also illuminate why it seems wrong to prioritize

¹²⁶ Another possible answer to the challenge is that some human benefits are more than merely finitely many times larger than any benefit to (common) low-welfare species. This would break the symmetry given my assumption that the benefits for the high-welfare species are only finitely many times larger than human benefits. This answer is natural for those who believe that, even among human welfare goods, some “lexically superior” goods contribute more to one’s welfare than any amount of other goods (e.g. Dorsey 2009). However, these views about human wellbeing have highly implausible implications (see e.g. Schönherr 2018). We may also wonder if there might be goods for high-welfare species that are lexically prior to any of our goods, and whether we would be willing to assign absolute priority to them under this assumption.

large benefits to members of high-welfare species over any number of any benefits for humans.

4. The Argument from Fairness

The first argument is based on the value of *fairness*. As far as this value is concerned, it seems plausible that we should at least sometimes benefit the members of a given species if we can benefit them effectively enough: for some (arbitrarily large) sum of benefits S, if we never benefit the members of a species even though we can benefit so many of them in individually significant ways that the sum of their benefits exceeds S, we are treating them unfairly. Intuitively, low-welfare species are not an exception to this rule. If anything, fairness requires that we give their welfare *special* consideration to compensate for their lower natural capacity for welfare, so that it is *especially* unfair not to benefit them if we can do so effectively.¹²⁷ Either way, fairness seems to require that, if we can benefit them effectively enough, we benefit low-welfare species at least sometimes.

This fairness requirement is in tension with the view that no number of even the most significant benefits for low-welfare species can outweigh one human life. This view implies that *no matter* how effectively we can benefit low-welfare species, we should not benefit them when we could save a human life alternatively. In a world like ours in which we could virtually always use our resources to save human lives with statistical certainty (e.g. by donating to effective charities), this would mean that we should *never* benefit low-welfare species even if we could benefit them much more effectively than we actually can. The only exception might be decisions about how to use small amounts of resources that

¹²⁷ This fairness requirement could be spelled out in egalitarian or prioritarian terms. Alternatively, it could be argued that *compassion* requires that we give low-welfare species special consideration. Crisp (2003, 761) entertains the view that an impartial spectator might “feel compassion exactly in proportion to levels of overall welfare”.

do not suffice to help any human significantly, such as when we decide whether to use a minute of our time to help a bee that is trapped in a room. However, many small amounts of resources add up to a big amount with which a human *could* be benefited significantly. As long as we foresee that there will be sufficiently many situations in which we could decide to spend small amounts of resources on low-welfare species, it seems, therefore, that we should decide to always save the resources: at some point, these savings will add up to an amount that is sufficient to significantly benefit a human. Arguably, we should not even waste our time, then, to help bees escape our room.¹²⁸

Even though we are used to prioritizing humans over (what we take to be) low-welfare species, this seems to be an extreme form of discrimination. Compare this to the fairness concern that a broadly utilitarian approach might require us to systematically prioritize the lives of healthy people over the lives of disabled people when the latter tend to contain less welfare (e.g. John et al. 2017; Kamm 2015). Whether or not it would be unfair to prioritize healthy lives over equally many disabled lives with lower welfare, which is a valid concern in itself, it *obviously* would be unfair to prioritize one healthy life over *any* number of disabled lives. The fairness claim that I am making with regard to low-welfare species is analogous to this extremely weak claim: it would be unfair to prioritize one human life over *any* number of lives of low-welfare species.

While the discrimination against low-welfare species might seem less morally grave, it is important to see that there still is a fairness concern that we don't have when it comes to prioritizing human lives over human *headaches*: intuitively, we do not discriminate against the headache patients by prioritizing someone's life over the alleviation of their headaches. On the contrary, it seems that it would be unfair against the

¹²⁸ Admittedly, it is often not possible to save non-monetary resources in the straightforward way in which we can save money. However, it also seems plausible for non-monetary resources that, *in expectation*, we will have more of these resources available for some valuable uses if we don't occasionally spend small amounts of them on other things. In particular, it seems that, in expectation, we will have more time available to benefit humans if we decide to never spend any time on helping insects (except when this benefits humans).

person whose life is at stake if we prioritized the headache patients. These intuitions suggest that, in decisions in which human lives are at stake, we should not treat significant benefits for low-welfare species in the way that we treat the alleviation of headaches; even if the latter is irrelevant in such decisions, the former seem to remain a relevant moral factor.

5. The Argument from Empathy

The second argument can be considered an argument from *empathy*. Unlike the argument from fairness, this argument is not based on an inter-species welfare comparison but focuses on the perspective of members of the species with the smaller welfare range. The point is that from *their* perspective the largest goods attainable to them are significant even if they are much smaller than the goods that are attainable to us. Of course, this may not be their *actual* perspective. After all, low-welfare species might lack the cognitive capacities to contemplate the significance of their goods. But we can still ask which perspective it would be *appropriate* to have for an individual with the welfare structure of a low-welfare species. From this (perhaps only hypothetical) perspective, it seems that some benefits for members of low-welfare species are significant. Perhaps, for example, the joy that they can experience is much less intense or much shorter than the joy that we are able to experience. Still, if that is the most joy that they can experience, experiencing it would constitute a highlight in their lives that is worth aspiring to. Or consider their lives: from our perspective they may seem insignificant because they have, for example, only a tiny fraction of the length of our lives. But for them, their lives are all that they have, and preventing their lives from being cut short is a worthy goal.

None of that is to say that their capacity for welfare is not in fact much lower than our capacity for welfare, or that this difference is morally irrelevant. However, this change in perspective is supposed to show that the goods of low-welfare species can have a

significance that makes it inappropriate to consider them irrelevant whenever important goods for us are at stake. It would be *arrogant* and perhaps *disrespectful* to consider their largest goods irrelevant whenever our largest goods are at stake.

6. Comparing Relative Magnitudes

I have argued that we should reject principles that ask us to prioritize a single human life over any number of any benefits for members of low-welfare species. How should we decide, then, in which cases some number of benefits of one kind can outweigh benefits of another kind? The arguments from fairness and empathy suggest that, rather than only considering the *absolute* magnitudes of benefits, we need to consider how large benefits are *relative* to the beneficiaries' welfare ranges or some related welfare quantity that differs between the beneficiaries. That is, we need to put each benefit in relation to this welfare quantity of the beneficiary, such as the beneficiary's welfare range, and then compare the resulting *relative benefit magnitudes*. If a benefit has a sufficiently large relative magnitude, it is relevant even if its absolute magnitude is small. This would allow us to maintain that each individual has some welfare goods that are never irrelevant, and that we should prioritize the members of each species in at least some hypothetical circumstances in which we could benefit them effectively enough.

In relation to which welfare quantity should we consider benefits? Since my discussion has been in terms of species with different welfare ranges, it is natural to assume that we should take the welfare range of the beneficiary's species. Although this would give us the right results in most cases, I think it is clear, however, that species' welfare ranges only matter derivatively. At least, they matter only insofar as they correlate with the *individual's* welfare range. This correlation is usually strong, but an individual's welfare range can also significantly come apart from the welfare range of its species, as in the case

of some irremediable chronic diseases. Benefits to such individuals should not be considered irrelevant merely because they are small relative to the species' welfare range. As long as a benefit is large compared to the individual's restricted welfare range, it should be considered relevant.

It could be suggested, furthermore, that individuals' welfare ranges only matter insofar as they correlate with their *actual* welfare. This would mean that we need to put benefits in relation to the beneficiary's actual welfare, which may be much lower than the maximum level of the beneficiary's welfare range, and then compare *these* relative magnitudes. This is a natural suggestion for those who take the actual welfare of the beneficiaries to affect the relevance of benefits for reasons that are independent of the consideration of low-welfare species (e.g. Voorhoeve 2014).¹²⁹ As noted before, however, if this view is to allow some tiny benefits for individuals with low welfare to be relevant, such as in the case of benefits for low-welfare species, it also needs to accept that tiny benefits for humans with equally low welfare are relevant. When, additionally, their welfare *range* is very small, this is not implausible. For example, the alleviation of headaches should be considered relevant to the saving of a human life in the very unusual case where the headache patients have an extremely limited capacity for welfare (e.g. due to an extremely severe disability), so that the benefit of alleviating their headaches is significant compared to the welfare that they could ever realistically achieve. This case is analogous to the case of conferring significant benefits to members of low-welfare species which, as I have already argued, should be considered relevant to the saving of human lives. However, should we also consider headaches to be relevant to human lives if the headache patients have a normal human capacity for welfare and only coincidentally have a very low welfare?

¹²⁹ More precisely, Voorhoeve (2014) suggests giving more weight to benefits for beneficiaries who *would* have a lower level of well-being if they are *not* benefited. Whether we take the beneficiaries' wellbeing conditional on being benefited or conditional on not being benefited does not make a difference to my subsequent argument.

The answer to this question decides whether we should choose the welfare range or the actual welfare of the beneficiary as the relational factor.¹³⁰

Whether we take the welfare range or the actual welfare of beneficiaries, I think that a principle on which we can determine the relevance of benefits by comparing their relative magnitudes is promising. Such a principle would give us the same plausible results as Compare Absolute Magnitudes for common intra-species comparisons in which the beneficiaries are assumed to have similar welfare ranges and to be (if benefited) similarly well off. In that case, the relational factor is roughly the same for all beneficiaries and therefore does not make a difference. We could thereby maintain, for example, the intuitive verdict that headaches cannot outweigh human lives when ‘all else is equal’. At the same time, comparing relative magnitudes avoids the problems discussed above for Compare Absolute Magnitudes that arise in inter-species comparisons. By comparing relative magnitudes of benefits, we would not discriminate against species with smaller welfare ranges and would do justice to their distinct perspectives. Most importantly, we could explain why the largest attainable benefits for a species, whether for humans or for low-welfare species, are always relevant: these benefits have the largest possible relative magnitude.

It is tempting to conclude that we *only* need to compare relative magnitudes to determine whether a given benefit is relevant: for all benefits B_1 and B_2 , B_1 is relevant to B_2 if and only if the relative magnitude of B_1 is sufficiently large compared to the relative magnitude of B_2 . This simple principle, however, which we can refer to as “Compare

¹³⁰ A potential (third) solution would be to tie the relational factor to the notion of *fortune*, which is generally taken to be a relation between the individual’s wellbeing and some appropriate standard (e.g. McMahan 2002). The difficult question is what this standard is, where both the species’ and the individual’s potential or capacity for wellbeing are debated candidates (ibid.; Vallentyne 2005). Given these parallels, it could be suggested that this standard, whatever it is, is the same as the factor relative to which benefits need to be considered when testing for their relevance. A principle on which we should compare relative benefit magnitudes could then be interpreted as a principle on which we should compare the increases in fortune that the benefits at issue would result in.

Relative Magnitudes”, has a serious problem as well: minor benefits for humans would be irrelevant compared to significant benefits for low-welfare species. Assume, for example, that alleviating a human headache has the same benefit magnitude as saving the life of a bee while the welfare range (as well as the actual welfare) of the human patients is much larger than the welfare range (and the actual welfare) of the bee. The *relative* benefit magnitude of alleviating the headaches is then much *smaller* than the relative benefit magnitude of saving the bee’s life. Compare Relative Magnitudes would therefore consider the human headaches to be irrelevant, so that we should save the bee’s life rather than alleviate *any* number of headaches. But surely we should spare the whole human population from suffering a headache rather than save the life of a single bee. Compare *Absolute* Magnitudes gives us the right result here: per assumption, the absolute benefit magnitudes of alleviating a human headache and of saving a bee’s life are roughly the same. Consequently, the human headaches count as relevant and can thus outweigh the bee’s life if they are numerous enough.

My tentative proposal is therefore to combine Compare Absolute Magnitudes and Compare Relative Magnitudes as follows: each principle provides a *sufficient* condition for a benefit’s relevance, but it is only necessary that *one* of these two sufficient conditions is fulfilled. That is: a given benefit is relevant if and only if its absolute magnitude *or* its relative magnitude is sufficiently large. More precisely, for all benefits B_1 and B_2 , B_1 is relevant to B_2 if and only if the absolute magnitude of B_1 is sufficiently large compared to the absolute magnitude of B_2 *or* the relative magnitude of B_1 is sufficiently large compared to the relative magnitude of B_2 .¹³¹ This principle – *Compare Absolute and Relative Magnitudes* – gives us the intuitively right results in all considered cases: headaches cannot

¹³¹ As with Compare Absolute Magnitudes, we could alter this principle to give extra weight to benefits for beneficiaries who are less well off (if we choose welfare ranges as the relational factor), and we could add further necessary conditions to, for example, exclude benefits that are based on immoral preferences.

outweigh human lives (since both the absolute and the relative magnitude of headaches is too small); headaches can outweigh bee lives (since the absolute magnitude of headaches is large enough); human lives can outweigh lives of high-welfare species (since the relative magnitude of human lives is large enough); and lives of low-welfare species can outweigh human lives (since the relative magnitude of lives of low-welfare species is large enough).

7. Conclusion

There is of course more to be said about Compare Absolute and Relative Magnitudes. For one thing, it is less parsimonious than the simple Compare Absolute Magnitudes and Compare Relative Magnitudes. However, I think that my discussion shows that there is at least a promising (type of) principle that can explain why some benefits for members of low-welfare species are relevant even when human lives are at stake. Since this verdict gains plausibility from considerations of fairness and empathy, while principles that deny it face the problem of high-welfare species, I conclude that the overall most plausible view is that some benefits for low-welfare species can in principle outweigh even the most significant benefits for humans. This conclusion is very preliminary, though. The arguments from empathy and from fairness are by no means knock-down arguments, and perhaps there is a principle that can respond to the challenge of high-welfare species in a satisfactory way. Further investigation of these issues is needed to reach a conclusion in which we can be more confident.

References

- Crisp, Roger. 2003. "Equality, Priority, and Compassion." *Ethics* 113 (4): 745–763.
- Dorsey, Dale. 2009. "Headaches, Lives and Value." *Utilitas* 21 (1): 36–58.

- Fischer, Bob. 2023. "Rethink Priorities' Welfare Range Estimates." *Rethink Priorities*, January 23. <https://rethinkpriorities.org/publications/welfare-range-estimates>.
- . 2024. *Weighing Animal Welfare: Comparing Well-being Across Species*. New York: Oxford University Press.
- Horta, Oscar, and Dayron Teran. 2023. "Reducing Wild Animal Suffering Effectively." *Ethics, Policy and Environment* 26 (2): 217–230.
- Horton, Joe. 2018. "Always aggregate." *Philosophy & Public Affairs* 46 (2): 160–174.
- John, Tyler M., Joseph Millum, and David Wasserman. 2017. "How to allocate scarce health resources without discriminating against people with disabilities." *Economics and Philosophy* 33 (2): 161–186. doi:10.1017/S0266267116000237.
- Kagan, Shelly. 2019. *How to Count Animals, more or less*. Oxford: Oxford University Press.
- Kamm, Frances. 2015. "Cost effectiveness analysis and fairness." *Journal of Practical Ethics* 3 (1).
- McMahan, Jeff. 2002. *The Ethics of Killing*. New York: Oxford University Press.
- Norcross, Alastair. 1997. "Comparing Harms: Headaches and Human Lives." *Philosophy & Public Affairs* 26 (2): 135–167.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Parfit, Derek. 1978. "Innumerate Ethics." *Philosophy & Public Affairs* 7 (4): 285–301.
- . 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Scanlon, Thomas M. 1998. *What We Owe to Each Other*. Cambridge, MA: Belknap Press of Harvard University Press.
- Schönherr, Julius. 2018. "Still Lives for Headaches: A Reply to Dorsey and Voorhoeve." *Utilitas* 30 (2): 209–218. <https://doi.org/10.1017/S0953820817000152>.
- Sebo, Jeff. 2023. "The Rebugnant Conclusion: Utilitarianism, Insects, Microbes, and AI Systems." *Ethics, Policy & Environment* 26 (2): 249–264. doi: 10.1080/21550085.2023.2200724.
- Shulman, Carl, and Nick Bostrom. 2021. "Sharing the World with Digital Minds." In *Rethinking Moral Status*, edited by Steve Clarke, Hazem Zohny, and Julian Savulescu. Oxford University Press. doi: 10.1093/oso/9780192894076.003.0018.
- Taurek, John M. 1977. "Should the Numbers Count?" *Philosophy & Public Affairs* 6: 293–316.

Vallentyne, Peter. 2005. "Of mice and men: Equality and Animals." *The Journal of Ethics* 9: 403–433. doi: 10.1007/s10892-005-3509-x.

Voorhoeve, Alex. 2014. "How Should We Aggregate Competing Claims?" *Ethics* 125 (1): 64–87.