Economic Theory (2020) 69:861–918
https://doi.org/10.1007/s00199-019-01194-8

RESEARCH ARTICLE

Rank-additive population ethics

Marcus Pivato[1]

Received: 8 November 2018 / Accepted: 17 April 2019 / Published online: 9 May 2019

Abstract

The class of *rank-additive* social welfare orders (RA SWOs) includes rank-weighted utilitarian, generalized utilitarian, and rank-discounted generalized utilitarian rules; it is a flexible framework for population ethics. This paper axiomatically characterizes RA SWOs and studies their properties in two frameworks: the *actualist* framework (which only tracks the utilities of people who actually exist) and the *possibilist* framework (which also assigns zero utilities to people who don't exist). The axiomatizations and properties are quite different in the two frameworks. For example, actualist RA SWOs can simultaneously evade the Repugnant Conclusion and promote equality, whereas in the possibilist framework, there is a trade-off between these two desiderata. On the other hand, possibilist RA SWOs satisfy the Positive expansion and Negative expansion axioms, whereas the actualist ones don't.

Keywords Population ethics · Repugnant Conclusion · Additively separable · Rank-dependent · Utilitarian

JEL Classification D63 · D71

1 Introduction

Present-day social and economic policies will not only affect the quality of life of future generations but also affect the number of people who exist in these generations.

I thank the participants of the *8th Murat Sertel Workshop* (Caen), the *Workshop on population, ethics and economic policy* (Paris), and the conferences *Public Economic Theory 2018* (Huẽ), *Social Choice and Welfare 2018* (Seoul) and the Karlsruhe Institute of Technology for their comments. I am particularly grateful to Matt Adler, Geir Asheim, Marc Fleurbaey, Hendrik Rommeswinkel, Colin Rowat, Dean Spears, John Weymark, and Stéphane Zuber for their suggestions. Finally, I thank two referees for many constructive suggestions. This research was supported by Labex MME-DII (ANR11-LBX-0023-01) and CHOp (ANR-17-CE26-0003).

✉ Marcus Pivato
marcuspivato@gmail.com

[1] THEMA, Université de Cergy-Pontoise, UFR Economie et Gestion, 33, boulevard du Port,
95011 Cergy-Pontoise Cedex, France

Thus, policy makers face a trade-off between the sheer number of future people and their quality of life. *Population ethics* is the analysis of such trade-offs using tools from social choice theory and moral philosophy. It arose as a response to the *Repugnant Conclusion*, an ethical paradox first identified by Parfit (1984). Parfit noted that, under seemingly plausible normative hypotheses, we should prefer a future where a hundred trillion people lead wretched lives that are barely worth living, over a world where a much smaller number (say, ten billion) lead lives of much higher quality. This disturbing observation is not only a *reductio ad absurdum* of classical utilitarianism: it also afflicts a wide variety of other moral systems, particularly versions of welfarist consequentialism. A variety of solutions have been proposed, but none are entirely satisfactory. Recent surveys of this literature are Arrhenius et al. (2019) and Greaves (2017). For book-length treatments, see Ryberg and Tännsjö (2004), Blackorby et al. (2005), Arrhenius (2018), and Arrhenius and Bykvist (2019).

Cowen (2004) observed that the Repugnant Conclusion has a similar structure to the Saint Petersburg Paradox: in both cases, the paradox arises when a valuable thing is allowed to become minuscule in one "dimension", as long as it simultaneously grows huge along some other dimensions. Cowen proposed that such paradoxes could be avoided by insisting that the value of any single dimension be bounded. But he did not formalize this idea. Earlier and independently, Sider (1991) had proposed a rule of population ethics he called "geometrism", which avoids the Repugnant Conclusion through precisely the boundedness strategy suggested by Cowen. But Sider was well aware of geometrism's shortcomings (in particular, its anti-egalitarianism), and he introduced it only as a counterexample to a conjectured impossibility result, not as a serious alternative. More recently, Asheim and Zuber (2014, 2017) have studied and axiomatically characterized *rank-discounted generalized utilitarianism*; like Sider's geometrism, it avoids the Repugnant Conclusion via Cowen's boundedness strategy, but unlike geometrism, it is also inequality-averse.

In this paper, I will introduce and axiomatically characterize a family of population ethical theories which generalize both Sider (1991) and Asheim and Zuber (2014, 2017). To define this family, I need some terminology. A *social outcome* specifies both what people exist and what the lifetime utility of each person is. A *social welfare order* (SWO) is an ordering over social outcomes; it embodies not only ethical judgements about the trade-offs we must make between the lifetime utilities of different people, but also ethical judgements about trade-offs we must make between these lifetime utilities and overall population size. For example, a SWO might judge that it is better to have a relatively small population of relatively happy people, than to have a much larger population of less happy people.[1]

I will assume that lifetime utilities are measured on an absolute, interpersonally comparable scale, where a lifetime utility of zero is the lower limit for a life which is "worth living". If someone's lifetime utility is positive, then this means that, *for her*, it is better to exist than not to exist. But if her lifetime utility is negative, then this means that, *for her*, it would have been better to not exist at all. Note that the fact that a person's life is worth living *for her* does not necessarily imply that it is ethically

[1] The term "SWO" is somewhat misleading, because this ordering may encode ethical judgements over *population size* in addition to ethical judgements over *welfare*. But I use it because it is standard. In the literature on population ethics, such orderings are also called *population principles* or *population axiologies*.

better that she exist; it may be that adding a particularly unhappy life to an already populous world is not an ethical improvement, even if the person who lives that life still regards it as worth living, on the balance.[2]

I will consider two kinds of SWO in this paper. They differ in the precise information encoded in the social outcome. In a *possibilist* SWO, a social outcome assigns a lifetime utility to all people who *could possibly exist*. If someone does not *actually* exist, then she is simply assigned a lifetime utility of zero in this representation. Thus, possibilist SWOs do not distinguish between an outcome where Alice exists but has a lifetime utility of zero (i.e. a life so wretched that she is indifferent to not existing), and an otherwise identical outcome where Alice simply doesn't exist at all. In contrast, in an *actualist* SWO, each social outcome specifies precisely which people exist. Thus, a clear distinction is made between an outcome where Alice exists but has a lifetime utility of zero, and an outcome where she doesn't exist.

In both the actualist and possibilist frameworks, I will investigate a family of SWOs that I call *rank-additive*. These are SWOs which admit an additively separable representation, like the classical utilitarian or prioritarian SWOs. Each lifetime utility is transformed by a continuous increasing function before summation. However, people are ranked in order from lowest to highest lifetime utility, and different transformations can be applied to different entries in this ranking. Thus, the person with the *highest* lifetime utility may have her utility transformed in a different way than a person with a lower lifetime utility, before summation. This generalizes *rank-weighted utilitarian* (or *generalized Gini*) social welfare orders (Donaldson and Weymark 1980; Weymark 1981; Yaari 1988; Bossert 1990). But like Ebert (1988) and Zank (2007), it allows different utility transformation functions (as opposed to merely different multiplicative weights) to be applied at different positions in the ranking.[3]

In defining rank-additive SWOs, there is a key difference between the actualist and possibilist frameworks. In an *actualist* SWO, we only rank, transform, and sum the lifetime utilities of the (finite) set of people who actually exist. By contrast, in a *possibilist* SWO, we rank, transform, and sum the lifetime utilities of everyone who could *possibly* exist—this includes a finite collection of nonzero utilities (amongst those who actually exist) and also an infinite collection of zero utilities (of those who do *not* exist). These zero utilities contribute nothing to the sum itself, but they have implications for how we rank the utilities of the people who *do* exist. Because of this, rank-additive axiologists have different functional forms in the actualist and possibilist frameworks, and admit different axiomatic characterizations. In particular, possibilist rank-additive SWOs always satisfy the axioms of Positive expansion and Negative expansion, which say that it is *always* good to add another person whose lifetime utility is above zero, and *never* good to add another person whose lifetime utility is

[2] In these assessments, it is important that we work with *lifetime utilities*, not momentary utilities. Thus, a judgement that "It would be ethically better if Alice did not exist" does not imply that Alice should die—rather, it means it would have been better if Alice had never been born. Now that Alice *does* exist, the axiological ordering will be increasing with respect to her lifetime utility, which in turn is typically an increasing function of her lifespan.

[3] Donaldson and Weymark (1980), Ebert (1988), and Bossert (1990) allow population size to vary, and impose consistency conditions between evaluations for different populations sizes. But they do not discuss population ethics; all comparisons in these papers involve two social outcomes with the *same* population, rather than two social outcomes with *different* populations.

below zero. This means they evade the *Sadistic Conclusion*, a paradox which afflicts critical-level utilitarianism, average utilitarianism, and many other proposed solutions to the Repugnant Conclusion (Arrhenius 2000). By contrast, actualist rank-additive SWOs almost never satisfy Positive expansion and Negative expansion, and hence frequently lead to the Sadistic Conclusion. On the other hand, actualist rank-additive SWOs easily reconcile inequality aversion with avoidance of the Repugnant Conclusion, whereas possibilist rank-additive SWOs do not. Sider's (1991) geometrism is a *possibilist* rank-additive SWO. Asheim and Zuber's (2014, 2017) rank-discounted utilitarianism is an *actualist* rank-additive SWO.

Most of the literature in population ethics adopts the actualist framework (e.g. Blackorby et al. 2005). Perhaps this is because of the suspicion that there is something nonsensical about imputing a utility to someone in a scenario where she does not even exist, or making welfare comparisons between scenarios where she exists and scenarios where she doesn't. But several authors have argued convincingly that one *can* make such welfare comparisons, once they are construed in the right way (Holtug 2001; Roberts 2003, §4; Adler 2008, §III.A; Adler 2019, §II.A; Arrhenius and Rabinowicz 2010, 2015; Fleurbaey and Voorhoeve 2015, §3). So possibilism cannot simply be rejected as logically incoherent. The choice between the possibilist and actualist frameworks thus turns on which of them offers more attractive solutions to the central problems of population ethics. As we shall see, each framework has advantages and disadvantages.

The remainder of the paper is organized as follows. Section 2 concerns possibilist SWOs. Section 2.1 introduces the formal framework and key examples. Section 2.2 contains the first main result of the paper: an axiomatic characterization of possibilist rank-additive SWOs. Section 2.3 contains further results, such as necessary and sufficient conditions for these SWOs to be inequality-averse and to evade the Repugnant Conclusion. Section 3 concerns actualist SWOs and has a similar structure: Sect. 3.1 introduces the framework and key examples, while Sect. 3.2 contains the second main result of the paper: an axiomatic characterization of actualist rank-additive SWOs. Section 3.3 contains further results. Section 4 discusses a major problem confronting all rank-additive SWOs—their violation of the axiom of *Existence independence*—and proposes some ways of mitigating this problem. Finally, Sect. 5 discusses some undesirable properties of rank-additive SWOs.

2 Possibilist social welfare orders

2.1 Definitions and examples

Let $\mathcal{I}$ be an infinite set, whose elements represent all the people who could ever exist. Let $\mathbb{R}^{\mathcal{I}}$ be the set of all infinite $\mathcal{I}$-indexed profiles $\mathbf{r} = (r_i)_{i\in\mathcal{I}}$ of real numbers. For all $i \in \mathcal{I}$, interpret r_i as the *lifetime utility* of individual i. If $r_i > 0$, then overall, i has a life worth living. If $r_i < 0$, then overall, i has a life *not* worth living—it would have been better if she had never existed at all. If $r_i = 0$, then i's life is indifferent to non-existence. This is usually referred to as the *neutral level* of lifetime utility. We will also set $r_i = 0$ in any scenario where i does *not* exist; the possibilist framework does not distinguish between non-existence and existence with a neutral lifetime utility.

Let $\mathcal{X}$ be the set of all elements of $\mathbb{R}^{\mathcal{I}}$ with only finitely many nonzero entries. An element of $\mathcal{X}$ represents a complete specification of all the lifetime utilities of all the people who will ever exist. (I assume this number to be finite.) I will refer to elements of $\mathcal{X}$ as *social outcomes*. A *possibilist social welfare order* is a preference order (i.e. complete, transitive, reflexive binary relation) $\succeq$ on $\mathcal{X}$. Let $\approx$ denote the symmetric part of $\succeq$, and let $\succ$ denote its asymmetric part.

If $\pi : \mathcal{I} \longrightarrow \mathcal{I}$ is any bijection, then define $\pi^* : \mathbb{R}^{\mathcal{I}} \longrightarrow \mathbb{R}^{\mathcal{I}}$ by setting $\pi^*(\mathbf{r}) := (r_{\pi(i)})_{i \in \mathcal{I}}$ for all $\mathbf{r} = (r_i)_{i \in \mathcal{I}}$ in $\mathbb{R}^{\mathcal{I}}$. Clearly, $\pi(\mathcal{X}) = \mathcal{X}$, and π restricted to $\mathcal{X}$ defines a bijection from $\mathcal{X}$ to itself. We will be interested in SWOs satisfying the following axiom:

Anonymity. If $\pi : \mathcal{I} \longrightarrow \mathcal{I}$ is any bijection, and $\mathbf{x} \in \mathcal{X}$, then $\mathbf{x} \approx \pi^*(\mathbf{x})$.

This is a standard axiom, which says that the SWO must treat all people the same. Let $\mathbb{R}_+ := \{r \in \mathbb{R};\ r \geqslant 0\}$ and let $\mathbb{R}_- := \{r \in \mathbb{R};\ r \leqslant 0\}$. Let $\mathbb{R}_+^\infty$ be the set of all infinite sequences $\mathbf{r} = (r_n)_{n=1}^\infty$ of nonnegative numbers. Let $\mathbb{R}_+^\propto$ be the set of all elements of $\mathbb{R}_+^\infty$ with only finitely many nonzero entries, and let $\mathbb{R}_+^{\propto\downarrow}$ be the set of all nonincreasing sequences in $\mathbb{R}_+^\propto$. Likewise define $\mathbb{R}_-^\infty$ and $\mathbb{R}_-^\propto$, and let $\mathbb{R}_-^{\propto\uparrow}$ be the set of all nondecreasing sequences in $\mathbb{R}_-^\propto$. For any $\mathbf{x} \in \mathcal{X}$, let $x_1^+ \geqslant x_2^+ \geqslant x_3^+ \geqslant \cdots \geqslant x_N^+ > 0$ be all the positive entries of $\mathbf{x}$, listed in decreasing order with each value appearing as many times in this list as it appears in $\mathbf{x}$, and define $\mathbf{x}^+ := (x_1^+, x_2^+, x_3^+, \ldots, x_N^+, 0, 0, \ldots)$, an element of $\mathbb{R}_+^{\propto\downarrow}$. Likewise, let $x_1^- \leqslant x_2^- \leqslant x_3^- \leqslant \cdots \leqslant x_N^- < 0$ be all the negative entries of $\mathbf{x}$, listed in increasing order with each value appearing as many times in this list as it appears in $\mathbf{x}$, and define $\mathbf{x}^- := (x_1^-, x_2^-, x_3^-, \ldots, x_M^-, 0, 0, \ldots)$, an element of $\mathbb{R}_-^{\propto\uparrow}$. Now define the function $\phi : \mathcal{X} \longrightarrow \mathbb{R}_+^{\propto\downarrow} \times \mathbb{R}_-^{\propto\uparrow}$ by setting $\phi(\mathbf{x}) := (\mathbf{x}^+, \mathbf{x}^-)$, for any $\mathbf{x} \in \mathcal{X}$. Clearly, ϕ is a surjection. If $\succeq_*$ is any preference order on $\mathbb{R}_+^{\propto\downarrow} \times \mathbb{R}_-^{\propto\uparrow}$, then we can define an SWO $\succeq$ on $\mathcal{X}$ by the formula:

$$\text{for all } \mathbf{x}, \mathbf{y} \in \mathcal{X} \qquad \big(\mathbf{x} \succeq \mathbf{y}\big) \iff \big(\phi(\mathbf{x}) \succeq_* \phi(\mathbf{y})\big). \tag{2A}$$

It is easy to see that $\succeq$ satisfies Anonymity: if $\pi : \mathcal{I} \longrightarrow \mathcal{I}$ is any bijection, and $\mathbf{x}' = \pi^*(\mathbf{x})$, then $\phi(\mathbf{x}') = \phi(\mathbf{x})$, so that $\mathbf{x} \approx \mathbf{x}'$. Conversely, if $\succeq$ is an SWO on $\mathcal{X}$ satisfying Anonymity, then there is a unique preference order $\succeq_*$ on $\mathbb{R}_+^{\propto\downarrow} \times \mathbb{R}_-^{\propto\uparrow}$ satisfying formula (2A). In other words, there is a natural bijective correspondence between preference orders on $\mathbb{R}_+^{\propto\downarrow} \times \mathbb{R}_-^{\propto\uparrow}$ and SWOs on $\mathcal{X}$ satisfying Anonymity.

A preference order on $\mathbb{R}_+^{\propto\downarrow} \times \mathbb{R}_-^{\propto\uparrow}$ enables us to treat a person's lifetime utility differently depending on how it is ranked relative to the lifetime utilities of other people. There are at least two reasons why this is useful. The first is aversion to inequality, which suggests that a small increase in the utility of a less happy person is more valuable than a similar increase in the utility of a more happy person, *ceteris paribus*. The second reason is specific to population ethics: the introduction of a new person with a relatively good life is generally regarded as desirable, whereas the introduction of a new person with a relatively mediocre life might be regarded as *un*desirable in some situations, even if that life is on the balance worth living. Both of these intuitions can be formalized using orderings on $\mathbb{R}_+^{\propto\downarrow} \times \mathbb{R}_-^{\propto\uparrow}$.

A *social welfare function*[4] (SWF) is a function $W : \mathcal{X} \longrightarrow \mathbb{R}$. It is *rank-additive* (RA) if there are continuous, increasing functions $\phi_n^+ : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ and $\phi_n^- : \mathbb{R}_- \longrightarrow \mathbb{R}_-$ with $\phi_n^+(0) = 0 = \phi_n^-(0)$ for all $n \in \mathbb{N}$, such that for any $\mathbf{x} \in \mathcal{X}$, we have

$$W(\mathbf{x}) = \sum_{n=1}^{\infty} \phi_n^+(x_n^+) + \sum_{n=1}^{\infty} \phi_n^-(x_n^-). \tag{2B}$$

(There are only finitely many nonzero summands, by the definition of $\mathcal{X}$.) A SWO $\succeq$ is *rank-additive* if it is represented by a rank-additive social welfare function. For example:

- Suppose that $\phi_n^{\pm}(r) = r$ for all $r \in \mathbb{R}_{\pm}$ and all $n \in \mathbb{N}$.[5] Then we obtain the *classical utilitarian* SWF, defined by

$$W(\mathbf{x}) = \sum_{n=1}^{\infty} x_n^+ + \sum_{n=1}^{\infty} x_n^- = \sum_{i \in \mathcal{I}} x_i, \qquad \text{for all } \mathbf{x} \in \mathcal{X}. \tag{2C}$$

- Let $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ be a continuous, increasing function with $\phi(0) = 0$. Suppose that $\phi_n^{\pm}(r) = \phi(r)$ for all $r \in \mathbb{R}_{\pm}$ and all $n \in \mathbb{N}$. Then we obtain the *generalized utilitarian* SWF, defined by

$$W(\mathbf{x}) = \sum_{n=1}^{\infty} \phi(x_n^+) + \sum_{n=1}^{\infty} \phi(x_n^-) = \sum_{i \in \mathcal{I}} \phi(x_i), \qquad \text{for all } \mathbf{x} \in \mathcal{X}. \tag{2D}$$

  In particular, if ϕ is strictly concave, then (2D) is called a *prioritarian* SWF, and exhibits inequality aversion.

- Let $\{c_n^+\}_{n=1}^{\infty}$ and $\{c_n^-\}_{n=1}^{\infty}$ be two sequences of positive constants. Suppose that $\phi_n^{\pm}(r) = c_n^{\pm} r$ for all $r \in \mathbb{R}_{\pm}$ and all $n \in \mathbb{N}$. Then we obtain the *rank-weighted utilitarian* SWF (Weymark 1981; Yaari 1988):

$$W(\mathbf{x}) = \sum_{n=1}^{\infty} c_n^+ x_n^+ + \sum_{n=1}^{\infty} c_n^- x_n^-, \qquad \text{for all } \mathbf{x} \in \mathcal{X}. \tag{2E}$$

- In particular, let $\beta \in (0, 1)$, and suppose that $\phi_n^{\pm}(r) = \beta^n r$ for all $r \in \mathbb{R}_{\pm}$ and all $n \in \mathbb{N}$. Then we obtain the *geometric* SWF proposed by Sider (1991):

$$W(\mathbf{x}) = \sum_{n=1}^{\infty} \beta^n x_n^+ + \sum_{n=1}^{\infty} \beta^n x_n^-, \qquad \text{for all } \mathbf{x} \in \mathcal{X}. \tag{2F}$$

[4] As noted in footnote 1, this terminology is somewhat misleading. But I use it because it is standard.

[5] In other words, $\phi_n^+(r) = r$ for all $r \in \mathbb{R}_+$ and $\phi_n^-(r) = r$ for all $r \in \mathbb{R}_-$. I will often use the notation "±" in this way to simultaneously make two assertions: one in which all uses of "±" in a particular statement become "+", and the other in which all uses of "±" in that statement become "−".

The classical utilitarian SWF (2C) arises as a special case of generalized utilitarianism (with $\phi(x) = x$) and rank-weighted utilitarianism (with $c_n^{\pm} = 1$ for all $n \in \mathbb{N}$). Unfortunately, as is well known, any generalized utilitarian SWF (and in particular, the classical utilitarian SWF) leads to Parfit's Repugnant Conclusion. In contrast, the rank-weighted utilitarian SWF (2E) evades the Repugnant Conclusion, as long as the sequence $\{c_n^+\}_{n=1}^{\infty}$ decays quickly enough that $\sum_{n=1}^{\infty} c_n^+ < \infty$ (see Proposition 2.2(a)). However, in this case, the rank-weighted utilitarian SWF is *anti*-egalitarian amongst all people with positive lifetime utility (see Proposition 2.4). To reconcile inequality aversion with evasion of the Repugnant Conclusion, we will need to consider rank-additive SWOs defined by other choices of functions $\{\phi_n^{\pm}\}_{n=1}^{\infty}$.

Rank-additive SWOs have several attractive properties. For any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ and $\mathbf{z} \in \mathbb{R}^N$, write "$\mathbf{y} = \mathbf{x} \uplus \mathbf{z}$" if there exist distinct $j_1, j_2, \ldots, j_N \in \mathcal{I}$ such that $x_{j_n} = 0$ and $y_{j_n} = z_n$ for all $n \in [1 \ldots N]$, while $x_i = y_i$ for all $i \in \mathcal{I} \setminus \{j_1, j_2, \ldots, j_N\}$.[6] In other words, $\mathbf{y}$ is obtained by adding to $\mathbf{x}$ exactly N new people, whose lifetime utilities are given by $(z_1, \ldots, z_N)$. For any $r \in \mathbb{R}$, we define $\mathbf{x} \uplus r := \mathbf{x} \uplus \mathbf{z}$, where $\mathbf{z}$ is an outcome containing a single individual with lifetime utility r. It is easily verified that any rank-additive possibilist SWO satisfies the next four axioms.

> Pareto. For all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, if $x_i \geqslant y_i$ for all $i \in \mathcal{I}$, then $\mathbf{x} \succeq \mathbf{y}$. If, furthermore, $x_i > y_i$ for some $i \in \mathcal{I}$, then $\mathbf{x} \succ \mathbf{y}$.
>
> Positive expansion (or Mere Addition). For any $\mathbf{x} \in \mathcal{X}$ and any $r > 0$, $\mathbf{x} \uplus r \succ \mathbf{x}$.
>
> Negative expansion. For any $\mathbf{x} \in \mathcal{X}$ and any $r < 0$, $\mathbf{x} \uplus r \prec \mathbf{x}$.
>
> No Sadistic Conclusion. For any $\mathbf{x} \in \mathcal{X}$, any $N, M \in \mathbb{N}$, and any $\mathbf{y} \in \mathbb{R}_{++}^N$ and $\mathbf{z} \in \mathbb{R}_{--}^M$, $\mathbf{x} \uplus \mathbf{y} \succ \mathbf{x} \uplus \mathbf{z}$.

Positive expansion says it is always good to add another person whose life is worth living (i.e. whose lifetime utility is positive). Negative expansion says it is always bad to add another person whose life is *not* worth living (i.e. whose lifetime utility is negative). Both of these are consequences of the Pareto axiom. Meanwhile, No Sadistic Conclusion is a consequence of Positive expansion and Negative expansion; it means that rank-additive possibilist SWOs avoid a well-known problem of average utilitarian and critical-level generalized utilitarian principles first identified by Arrhenius (2000).[7] For any $\mathbf{x} \in \mathbb{R}_{\pm}^{\infty}$, let $|\mathbf{x}|$ denote the number of nonzero entries in $\mathbf{x}$. Rank-additive possibilist SWOs also satisfy the next axiom.

> Existence independence of the wretched. For all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ and $N, M \in \mathbb{N}$ such that $|\mathbf{x}^+| = |\mathbf{y}^+| = N$ and $|\mathbf{x}^-| = |\mathbf{y}^-| = M$, and any $z \in \mathbb{R}$ such that $\max\{x_M^-, y_M^-\} \leqslant z \leqslant \min\{x_N^+, y_N^+\}$, we have $\mathbf{x} \succeq \mathbf{y}$ if and only if $\mathbf{x} \uplus z \succeq \mathbf{y} \uplus z$.

This axiom is similar to the axiom of *Existence independence* of Blackorby et al. (2005, §5.6), but it only applies the people whose lifetime utilities are close to zero ("the wretched").[8]

An important feature of RA possibilist SWOs is that individuals with positive lifetime utilities (i.e. lives worth living) are evaluated using the functions $\{\phi_n^+\}_{n=1}^{\infty}$,

[6] Throughout this document, the notation "$[1 \ldots N]$" refers to the set $\{1, \ldots, N\}$. Likewise, "$[N \ldots \infty)$" refers to the set $\{N, N+1, \ldots\}$.

[7] See just formula (3C) for the definition of critical-level utilitarianism.

[8] See Sect. 4 for further discussion of the *Existence independence* axiom.

whereas individuals with negative lifetime utilities (i.e. lives *not* worth living) are evaluated using $\{\phi_n^-\}_{n=1}^\infty$. This gives us the freedom to treat lives which are not worth living in a completely different way than we treat lives worth living, in accord with many people's ethical intuitions. For example, if we augment a social outcome by adding a trillion wretched lives that are barely worth living, then a rejection of the Repugnant Conclusion suggests that the marginal gain in social welfare obtained by adding the trillionth such life is less than the marginal gain from adding the first such life. But if we add a trillion lives of terrible suffering that are clearly *not* worth living, then our ethical intuitions suggest that the addition of the trillionth such life adds just as much evil to the world as the first one. This intuition is sometimes called *the asymmetry* (McMahan 1981, 2009; Roberts 2011, 2019). Since $\{\phi_n^+\}_{n=1}^\infty$ and $\{\phi_n^-\}_{n=1}^\infty$ can have different properties, it is easy to accommodate this intuition.

It seems natural to assume that $\{\phi_n^+\}_{n=1}^\infty$ and $\{\phi_n^-\}_{n=1}^\infty$ should both arise as restrictions to $\mathbb{R}_+$ and $\mathbb{R}_-$ of some common family of utility functions defined on all of $\mathbb{R}$, as in the generalized utilitarian SWF in formula (2D). If they didn't, and we treated negative and positive utilities in a completely different way, then one might worry about creating an "ethical discontinuity" in our treatment of an individual as her lifetime utility changes from positive to negative. But this concern is misconceived. To understand this, let $\mathbf{x} \in \mathcal{X}$ be a social outcome, and define $\mathbf{x}^+ = (x_1^+, x_2^+, x_3^+, \ldots, x_N^+, 0, 0, \ldots)$ and $\mathbf{x}^- = (x_1^-, x_2^-, x_3^-, \ldots, x_M^-, 0, 0, \ldots)$ as prior to statement (2A). If we imagine *all* the coordinates of $\mathbf{x}$ of being arranged in decreasing order, then the (infinite) number of zero coordinates will all appear *between* the coordinates of $\mathbf{x}^+$ and those of $\mathbf{x}^-$. In other words,

$$\mathbf{x} \;=\; (x_1^+, x_2^+, x_3^+, \ldots, x_N^+, 0, 0, 0, \ldots\ldots\ \ldots\ldots, 0, 0, 0, x_M^-, \ldots, x_3^-, x_2^-, x_1^-).$$

Observe that $\{\phi_n^+\}_{n=1}^\infty$ deal with the coordinates at the left end of this infinite array, while $\{\phi_n^-\}_{n=1}^\infty$ deal with the coordinates at the right end. There is no reason to believe that these two families of functions should have anything in common with one another. Indeed, suppose we gradually reduce one individual's lifetime utility, while holding all other utilities constant. As her utility decreases, it is shuffled further and further rightward in the ordering of $x_1^+, \ldots, x_N^+$. But when it passes from positive to negative, it jumps an *infinite* number of positions rightward (leaping over the infinite number of zero coordinates), to become part of $x_M^-, \ldots, x_1^-$. If there is an "ethical discontinuity" in our treatment of the person at this moment, it can be attributed to this infinite jump.

2.2 Axiomatic characterization

The first main result of the paper is an axiomatic characterization of rank-additive SWOs in the possibilist framework. This will use the Anonymity and Pareto axioms introduced in Sect. 2.1, along with two other axioms. For any $N \in \mathbb{N}$, let $\mathbb{R}_+^{N\downarrow} := \{\mathbf{r} \in \mathbb{R}^N;\ r_1 \geqslant r_2 \geqslant \cdots \geqslant r_N \geqslant 0\}$, and let $\mathbb{R}_-^{N\uparrow} := \{\mathbf{r} \in \mathbb{R}^N;\ r_1 \leqslant r_2 \leqslant \cdots \leqslant r_N \leqslant 0\}$. We can treat $\mathbb{R}_+^{N\downarrow}$ as a subset of $\mathbb{R}_+^{\infty\downarrow}$ in a natural way, by identifying the N-tuple $(x_1, x_2, \ldots, x_N)$ with the sequence $(x_1, x_2, \ldots, x_N, 0, 0, \ldots)$. Likewise, we can treat $\mathbb{R}_-^{N\uparrow}$ as a subset of $\mathbb{R}_-^{\propto\uparrow}$. Note that $\mathbb{R}_+^{2\downarrow} \subset \mathbb{R}_+^{3\downarrow} \subset \mathbb{R}_+^{4\downarrow} \subset \cdots \subset \mathbb{R}_+^{\propto\downarrow}$ and $\mathbb{R}_-^{2\uparrow} \subset \mathbb{R}_-^{3\uparrow} \subset \mathbb{R}_-^{4\uparrow} \subset \cdots \subset \mathbb{R}_-^{\propto\uparrow}$. It follows that[9]

$$\mathbb{R}_+^{\propto\downarrow} \times \mathbb{R}_-^{\propto\uparrow} = \bigcup_{N=1}^{\infty} \left(\mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow} \right). \tag{2G}$$

For all $N \in \mathbb{N}$, let $\succeq_N$ be the restriction of $\succeq_*$ to $\mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow}$. The order $\succeq_*$ is uniquely determined by this sequence $(\succeq_N)_{N=1}^{\infty}$ of finite-population SWOs. The next two axioms concern these orders. Note that $\mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow}$ is a closed convex subset of $\mathbb{R}^N \times \mathbb{R}^N = \mathbb{R}^{2N}$; endow it with the subspace topology it inherits from $\mathbb{R}^{2N}$. We need two more axioms.

> Continuity. For every $N \in \mathbb{N}$, the order $\succeq_N$ is continuous on $\mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow}$.
>
> Separability. For every $N \in \mathbb{N}$, and every pair of subsets $\mathcal{J}_+, \mathcal{J}_- \subseteq [1 \ldots N]$, there is a preference order $\succeq_{\mathcal{J}_\pm}$ defined on $\mathbb{R}^{\mathcal{J}_+} \times \mathbb{R}^{\mathcal{J}_-}$ such that, for any $\mathbf{x} = (\mathbf{x}^+, \mathbf{x}^-)$ and $\mathbf{y} = (\mathbf{y}^+, \mathbf{y}^-)$ in $\mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow}$, if $x_n^+ = y_n^+$ for all $n \in [1 \ldots N] \setminus \mathcal{J}_+$, and $x_n^- = y_n^-$ for all $n \in [1 \ldots N] \setminus \mathcal{J}_-$, then $\mathbf{x} \succeq_N \mathbf{y}$ if and only if $(\mathbf{x}_{\mathcal{J}_+}^+, \mathbf{x}_{\mathcal{J}_-}^-) \succeq_{\mathcal{J}_\pm} (\mathbf{y}_{\mathcal{J}_+}^+, \mathbf{y}_{\mathcal{J}_-}^-)$.[10]

These axioms are somewhat weaker than the familiar axioms with similar names: they only apply to the restriction of $\succeq$ to a population of fixed, finite size N, and only compare social outcomes that are comonotonic. The Continuity axiom says that a small change in the lifetime utilities of individuals should only cause a small change in the ranking of the social outcome, in comparison with other social outcomes having the same population. Separability says that, if certain people (namely those in $[1 \ldots N] \setminus \mathcal{J}_+$) have the same positive lifetime utility in two social outcomes $\mathbf{x}$ and $\mathbf{y}$, and furthermore occupy the same position in the ranking from best-off to worst-off, then the comparison between $\mathbf{x}$ and $\mathbf{y}$ should be entirely determined by *other* people—those whose utilities differ from $\mathbf{x}$ to $\mathbf{y}$ (namely those in $\mathcal{J}_+$). Likewise, if certain people (namely those in $[1 \ldots N] \setminus \mathcal{J}_-$) have the same negative lifetime utility in two social outcomes $\mathbf{x}$ and $\mathbf{y}$, and furthermore occupy the same position in the ranking from worst-off to best-off, then the comparison between $\mathbf{x}$ and $\mathbf{y}$ should be entirely determined by *other* people—those whose utilities differ from $\mathbf{x}$ to $\mathbf{y}$ (namely those in $\mathcal{J}_-$). Thus, this axiom has the same normative content as the standard *Separability* axiom—it is just more complicated to state (and in fact, logically weaker), because it must deal separately with people having positive and negative lifetime utilities, and furthermore only deals with rank-ordered vectors of utility. Here is the first main result of the paper.

[9] To see this, let $(\mathbf{x}^+, \mathbf{x}^-) \in \mathbb{R}_+^{\propto\downarrow} \times \mathbb{R}_-^{\propto\uparrow}$. Then $\mathbf{x}^+ \in \mathbb{R}_+^{P\downarrow}$ and $\mathbf{x}^- \in \mathbb{R}_-^{N\uparrow}$ for some $P, N \in \mathbb{N}$. Let $M := \max\{P, N\}$; then, $\mathbb{R}_+^{P\downarrow} \subseteq \mathbb{R}_+^{M\downarrow}$ and $\mathbb{R}_-^{N\uparrow} \subseteq \mathbb{R}_-^{M\uparrow}$, so that $(\mathbf{x}^+, \mathbf{x}^-) \in \mathbb{R}_+^{M\downarrow} \times \mathbb{R}_-^{M\uparrow}$.

[10] Here, $\mathbf{x}_{\mathcal{J}_\pm}^{\pm} = (x_j^{\pm})_{j \in \mathcal{J}_\pm}$, an element of $\mathbb{R}^{\mathcal{J}_\pm}$. Strictly speaking, the order $\succeq_{\mathcal{J}_\pm}$ need only be defined on $\mathbb{R}_+^{\mathcal{J}_\downarrow^+} \times \mathbb{R}_-^{\mathcal{J}_-^{\uparrow}}$. But it makes no difference if we suppose it is defined on all of $\mathbb{R}^{\mathcal{J}_+} \times \mathbb{R}^{\mathcal{J}_-}$.

Theorem 1 *Let $\succeq$ be a possibilist SWO on $\mathcal{X}$. Then $\succeq$ satisfies* Anonymity*,* Continuity*,* Pareto*, and* Separability *if and only if it is rank-additive. Furthermore, in the representation* (2B)*, the functions $\{\phi_n^{\pm}\}_{n=1}^{\infty}$ are unique up to multiplication by a common positive constant.*

Proof sketch For all $N \in \mathbb{N}$ and all $\mathbf{x} \in \mathbb{R}_+^{N_\downarrow} \times \mathbb{R}_-^{N_\uparrow}$, use the classic representation theorem of Debreu (1960) to obtain a "local" additive representation of $\succeq_N$ in a neighbourhood of $\mathbf{x}$. Then glue together these local representations to obtain a single, *global* additive representation on all of $\mathbb{R}_+^{N_\downarrow} \times \mathbb{R}_-^{N_\uparrow}$, using a result of Chateauneuf and Wakker (1993). Finally, for any $N < M$, use the natural embedding $\mathbb{R}_+^{N_\downarrow} \times \mathbb{R}_-^{N_\uparrow} \subset \mathbb{R}_+^{M_\downarrow} \times \mathbb{R}_-^{M_\uparrow}$ together with standard uniqueness results to show that the $2N$ utility functions involved in the additive representation on $\mathbb{R}_+^{N_\downarrow} \times \mathbb{R}_-^{N_\uparrow}$ are actually the first $2N$ utility functions involved in the additive representation on $\mathbb{R}_+^{M_\downarrow} \times \mathbb{R}_-^{M_\uparrow}$. See Appendix A for details.

Independence of the axioms Consider a rank-additive SWF (2B). If some of the functions $\{\phi_n^{\pm}\}_{n=1}^{\infty}$ are *not* increasing, but all are continuous, then the resulting SWO violates Pareto but satisfies the other three axioms of Theorem 1. If, on the other hand, some of $\{\phi_n^{\pm}\}_{n=1}^{\infty}$ are not continuous, but all are increasing, then the resulting SWO violates Continuity but satisfies the other three axioms.

Next, suppose we replace (2B) with the expression $\left(\sum_{n=1}^{\infty} x_n^+\right)^2 - \left(\sum_{n=1}^{\infty} |x_n^-|\right)^2$. The resulting SWO violates Separability but satisfies the other three axioms. Finally, for all $i \in \mathcal{I}$, let $\phi_i : \mathbb{R} \longrightarrow \mathbb{R}$ be a continuous, increasing function with $\phi_i(0) = 0$, and define $W : \mathcal{X} \longrightarrow \mathbb{R}$ by setting $W(\mathbf{x}) := \sum_{i \in \mathcal{I}} \phi_i(x_i)$ (well defined because only finitely many summands are nonzero). The resulting SWO $\succeq$ satisfies Pareto, but if the functions $\{\phi_i\}_{i \in \mathcal{I}}$ are distinct, then it violates Anonymity. It is meaningless to ask whether $\succeq$ satisfies Continuity or Separability, since these axioms are formulated in terms of the order $\succeq_*$, which is not even well defined if $\succeq$ violates Anonymity. But for any finite subset $\mathcal{J} \subset \mathcal{I}$, if we define $\succeq_{\mathcal{J}}$ to be restriction of $\succeq$ to the (finite-dimensional) subspace $\mathbb{R}^{\mathcal{J}} \times \{0\}^{\mathcal{I} \setminus \mathcal{J}}$, then it is easy to see that $\succeq_{\mathcal{J}}$ satisfies axioms analogous to Continuity and Separability.

2.3 Further results

I earlier noted that any rank-additive SWO satisfies the axiom Existence independence of the wretched. We might also consider SWOs that satisfy the following axioms:

Top-independence in good worlds. For all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ such that $x_i \geqslant 0$ and $y_i \geqslant 0$ for all $i \in \mathcal{I}$, and all $z \in \mathbb{R}$ with $z > \max\{x_1^+, y_1^+\}$, we have $\mathbf{x} \succeq \mathbf{y}$ if and only if $\mathbf{x} \uplus z \succeq \mathbf{y} \uplus z$.

Bottom-independence in bad worlds. For all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ such that $x_i \leqslant 0$ and $y_i \leqslant 0$ for all $i \in \mathcal{I}$, and all $z \in \mathbb{R}$ with $z < \max\{x_1^-, y_1^-\}$, we have $\mathbf{x} \succeq \mathbf{y}$ if and only if $\mathbf{x} \uplus z \succeq \mathbf{y} \uplus z$.

These axioms are similar to Separability; they say that, when comparing certain kinds of social outcomes, we can disregard the utility—or even the *existence*—of certain people who are indifferent between these outcomes. The first axiom is like Asheim and Zuber's (2014) axiom *Existence independence of the best off*, except that it applies only in "good" worlds, where everyone's lifetime utility is nonnegative. The second axiom is like Asheim and Zuber's (2014) *Existence independence of the worst off*, but it applies only in "bad" worlds.[11] The next result says that these axioms lead to something resembling Sider's (1991) "geometric" SWF from formula (2F).

Proposition 2.1 *Let $\succeq$ be a rank-additive possibilist SWO with the SWF* (2B).

(a) $\succeq$ *satisfies* Top-independence in good worlds *if and only if there is a continuous, increasing function* $\phi^+ : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ *(unique up to multiplication by a positive constant) and a unique constant* $\beta_+ > 0$ *such that* $\phi_n^+ = \beta_+^n \, \phi^+$ *for all* $n \in \mathbb{N}$.

(b) $\succeq$ *satisfies* Bottom-independence in bad worlds *if and only if there is a continuous, increasing function* $\phi^- : \mathbb{R}_- \longrightarrow \mathbb{R}_-$ *(unique up to multiplication by a positive constant) and a unique constant* $\beta_- > 0$ *such that* $\phi_n^- = \beta_-^n \, \phi^-$ *for all* $n \in \mathbb{N}$.

(c) *If* $\succeq$ *satisfies the conditions of both* (a) *and* (b), *then* ϕ^+ *and* ϕ^- *are unique up to multiplication by the same positive constant.*

If $\succeq$ *satisfies both* Top-independence in good worlds *and* Bottom-independence in bad worlds, *then Proposition* 2.1 *yields a variant of Sider's geometric SWF. But nothing in Proposition* 2.1 *requires* $\beta_+ = \beta_-$, *nor are either* β_+ *or* β_- *required be less than 1.*

Repugnant Conclusions For any $N \in \mathbb{N}$, let $\mathbf{1}_N$ refer to an element of $\mathcal{X}$ such that exactly N coordinates take the value 1, and all other coordinates are zero. (By Anonymity, it does not matter which coordinates we choose.) For any $r \in \mathbb{R}$, $r\,\mathbf{1}_N$ refers to the corresponding element of $\mathcal{X}$ such that exactly N coordinates take the value r, and all other coordinates are zero. Consider the following axioms.

> No Repugnant Conclusion. There exist $r_0 > 0$ and $\mathbf{x} \in \mathcal{X}$ such that $\mathbf{x} \succ r_0\,\mathbf{1}_N$ for all $N \in \mathbb{N}$.
>
> No utility monsters. For all $N \in \mathbb{N}$, there exists $\mathbf{x} \in \mathcal{X}$ such that $\mathbf{x} \succ r\,\mathbf{1}_N$ for all $r > 0$.

The first axiom rules out Parfit's Repugnant Conclusion. It says there is a minimum positive utility r_0 (representing a life which is technically worth living, but perhaps not very pleasant) and a social outcome $\mathbf{x}$ (e.g. the population of a modern industrialized country) which is better than *any* population of people with life utilities less than or equal to r_0, no matter how large this population becomes. The second axiom rules out Nozick's (1974) *Utility Monster* paradox. It says that for any finite-population size N, there exists a social outcome (presumably involving a larger number of people) which is better than *any* society which involves only N people, no matter how high their lifetime utilities becomes. Thus, even if the first N people are somehow much

[11] Asheim and Zuber's axioms are also stronger in that they compare populations of different sizes. Also similar are the axioms HIGAP and LIGAP, used by Bossert (1990) to characterize single-series Gini SWOs.

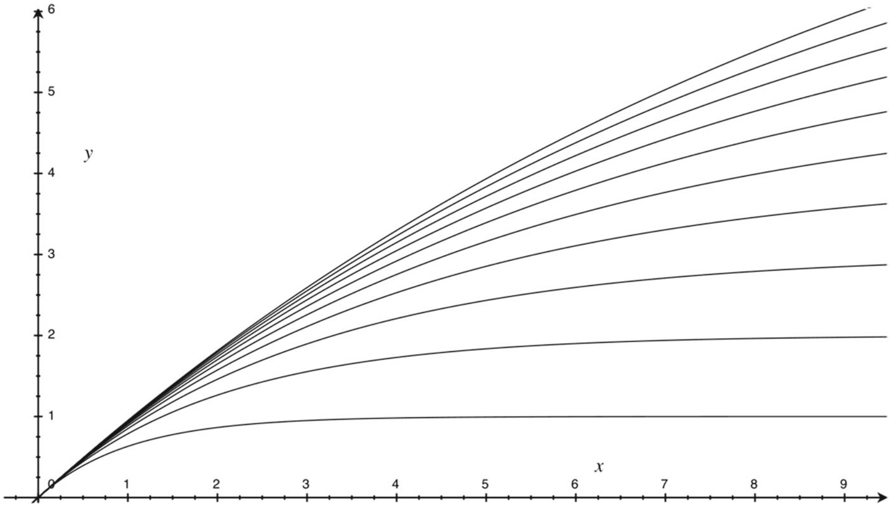

Fig. 1 Functions $\phi_n^+(r) = n \cdot (1 - \exp(-r/n))$, for $n \in \{1 \ldots 10\}$

more efficient at converting resources into lifetime utility than everyone else, the SWF does not allow them to simply absorb unlimited amounts of resources from the rest of humanity to boost their own utilities.

Proposition 2.2 *Let $\succeq$ be a rank-additive possibilist SWO with the SWF* (2B). *Let* $\overline{W} := \sup\{W(\mathbf{x});\ \mathbf{x} \in \mathcal{X}\}$.

(a) $\succeq$ *satisfies* No Repugnant Conclusion *if and only if there exists $r_0 > 0$ such that* $\displaystyle\sum_{n=1}^{\infty} \phi_n^+(r_0) < \overline{W}$.

(b) $\succeq$ *satisfies* No utility monsters *if and only if* $\lim_{r\to\infty} \sum_{n=1}^{N} \phi_n^+(r) < \overline{W}$, *for all* $N \in \mathbb{N}$.

If $\overline{W} < \infty$, *then both these conditions are satisfied.*

It is well known that Nozick's Utility Monster paradox can be evaded by using a generalized utilitarian social welfare like (2D) when the function ϕ is bounded above. In particular, some prioritarian social welfare functions have this form. But Proposition 2.2(b) goes beyond this trite observation, because in an RA SWF, the functions $\{\phi_n^+\}_{n=1}^{\infty}$ need not be identical, so they need not have the same upper bound. For example, suppose that $\phi_n^+(r) := n \cdot (1 - \exp(-r/n))$ for all $n \in \mathbb{N}$ and all $r \in \mathbb{R}_+$; then, the condition of Proposition 2.2(b) is satisfied. However, as shown in Fig. 1, we have $\sup(\phi_n^+(\mathbb{R}_+)) = n$ for all $n \in \mathbb{N}$.

The Saint Petersburg paradox The Repugnant Conclusion and the Utility Monster are both ethical paradoxes which arise when a valuable thing is allowed to become extremely small in one "dimension", as long as it simultaneously grows extremely large along some other dimensions. Perhaps the earliest paradox of this kind is the

Saint Petersburg Paradox (Bernoulli, 1738 [1954]). This suggests the next axiom. Let $W : \mathcal{X} \longrightarrow \mathbb{R}$ denote a SWF representing a SWO $\succeq$.

> No Saint Petersburg Paradox. There is some $\epsilon > 0$ and some $\mathbf{x} \in \mathcal{X}$ such that for any $\mathbf{y} \in \mathcal{X}$ and any $p < \epsilon$, $W(\mathbf{x}) > p\, W(\mathbf{y})$.

Let $\mathbf{z}$ be an outcome such that $W(\mathbf{z}) = 0$ (for example, the "empty world" where $z_i = 0$ for all $i \in \mathcal{I}$). We can interpret $p\, W(\mathbf{y})$ as the *expected W value* of a lottery which yields $\mathbf{y}$ with probability p, and $\mathbf{z}$ with probability $(1 - p)$. Thus, No Saint Petersburg Paradox says that the sure outcome $\mathbf{x}$ is better than any such lottery, no matter how good $\mathbf{y}$ is.[12]

Proposition 2.3 *Let $\succeq$ be a rank-additive possibilist SWO with the SWF (2B). Then $\succeq$ satisfies* No Saint Petersburg Paradox *if and only if* $\sup\{W(\mathbf{x});\ \mathbf{x} \in \mathcal{X}\} < \infty$. *In this case, $\succeq$ automatically satisfies* No Repugnant Conclusion *and* No utility monsters.

Inequality aversion Let $\mathbf{x}, \mathbf{y} \in \mathcal{X}$. Say that $\mathbf{y}$ is a *Pigou–Dalton transform* of $\mathbf{x}$ if there exist $j, k \in \mathcal{I}$ and $\epsilon > 0$ such that $y_j = x_j + \epsilon \leqslant y_k = x_k - \epsilon$, while $y_i = x_i$ for all other $i \in \mathcal{I} \setminus \{j, k\}$. Consider the following axioms.

> Inequality neutrality. Let $\mathbf{x}, \mathbf{y} \in \mathcal{X}$. If $\mathbf{y}$ is a Pigou–Dalton transform of $\mathbf{x}$, then $\mathbf{y} \approx \mathbf{x}$.
> Inequality aversion. Let $\mathbf{x}, \mathbf{y} \in \mathcal{X}$. If $\mathbf{y}$ is a Pigou–Dalton transform of $\mathbf{x}$, then $\mathbf{y} \succeq \mathbf{x}$.
> Strict inequality aversion. Let $\mathbf{x}, \mathbf{y} \in \mathcal{X}$. If $\mathbf{y}$ is a Pigou–Dalton transform of $\mathbf{x}$, then $\mathbf{y} \succ \mathbf{x}$.

Proposition 2.4 *Let $\succeq$ be a rank-additive possibilist SWO with the SWF* (2B).

(a) *$\succeq$ satisfies* Inequality neutrality *if and only if it is classical utilitarianism.*
(b) *$\succeq$ satisfies* Inequality aversion *if and only if, for all $n, m \in \mathbb{N}$, $r, s \in \mathbb{R}$ and $\epsilon > 0$:*

(i) *If $r \geqslant 0 \geqslant s$, then $\phi_n^+(r + \epsilon) - \phi_n^+(r) \leqslant \phi_m^-(s) - \phi_m^-(s - \epsilon)$.*
(ii) *If $n < m$ and $r \geqslant s \geqslant \epsilon > 0$, then $\phi_n^+(r + \epsilon) - \phi_n^+(r) \leqslant \phi_m^+(s) - \phi_m^+(s - \epsilon)$.*
(iii) *If $n > m$ and $s \leqslant r \leqslant -\epsilon < 0$, then $\phi_n^-(r + \epsilon) - \phi_n^-(r) \leqslant \phi_m^-(s) - \phi_m^-(s - \epsilon)$.*

Thus, for all $q \in \mathbb{R}_+$, we have

$$\begin{aligned} &\phi_1^+(q) \leqslant \phi_2^+(q) \leqslant \phi_3^+(q) \leqslant \cdots\cdots \\ &\quad \cdots\cdots \leqslant -\phi_3^-(-q) \leqslant -\phi_2^-(-q) \leqslant -\phi_1^-(-q). \end{aligned} \tag{2H}$$

In particular, if $\{\phi_n^+\}_{n=1}^\infty$ and $\{\phi_n^-\}_{n=1}^\infty$ are differentiable, then for any positive nonincreasing sequence $r_1 \geqslant r_2 \geqslant r_3 \geqslant \cdots \geqslant 0$ and negative nondecreasing sequence $s_1 \leqslant s_2 \leqslant s_3 \leqslant \cdots \leqslant 0$,

[12] Technically, this "lottery" interpretation goes beyond the formal framework of the rest of the paper, which involves no risk. But in reality, social decisions always involve risk. I have confined the analysis to riskless decisions only for simplicity. We could explicitly model risk using social lotteries, and then, No Saint Petersburg Paradox could be stated directly in terms of such lotteries. But this would also raise many other issues that are beyond the scope of this paper; see, for example, Mongin and Pivato (2015, 2016, 2018).

$$(\phi_1^+)'(r_1) \leqslant (\phi_2^+)'(r_2) \leqslant (\phi_3^+)'(r_3) \leqslant \cdots\cdots$$
$$\cdots\cdots \leqslant (\phi_3^-)'(s_3) \leqslant (\phi_2^-)'(s_2) \leqslant (\phi_1^-)'(s_1). \tag{2I}$$

(c) $\succeq$ *satisfies* Strict inequality aversion *if and only if all the statements in part* (b) *hold with strict inequalities.*

Example 2.5 The generalized utilitarian SWF (2D) satisfies Inequality aversion if and only if the function ϕ is concave; it satisfies Strict inequality aversion if and only if ϕ is strictly concave. The rank-weighted utilitarian SWF (2E) satisfies Inequality aversion if and only if $c_1^+ \leqslant c_2^+ \leqslant c_3^+ \leqslant \cdots \leqslant c_3^- \leqslant c_2^- \leqslant c_1^-$; it satisfies Strict inequality aversion if and only if these inequalities are all strict. Note, however, that in a general rank-additive SWF, we do not need the functions $\{\phi_n^+\}_{n=1}^\infty$ and $\{\phi_n^-\}_{n=1}^\infty$ to be concave to ensure inequality aversion, as long as the conditions of Proposition 2.4 are satisfied. □

Unfortunately, by comparing Proposition 2.4 with Proposition 2.2(a), one sees that it is impossible to simultaneously satisfy Inequality aversion and No Repugnant Conclusion. If $\epsilon > 0$ is small, then No Repugnant Conclusion requires the sequence $\{\phi_n^+(\epsilon)\}_{n=1}^\infty$ to be summable, whereas Inequality aversion requires this sequence to be nondecreasing as in (2H)—a contradiction. Thus, to avoid the Repugnant Conclusion, we must somehow weaken Inequality aversion. Let $\theta > 0$. Let $\mathbf{x}, \mathbf{y} \in \mathcal{X}$. Say that $\mathbf{y}$ is a *θ-restricted Pigou–Dalton transform* of $\mathbf{x}$ if there exist $j, k \in \mathcal{I}$ and $\epsilon > 0$ such that $y_j = x_j + \epsilon \leqslant y_k = x_k - \epsilon$, while $y_i = x_i$ for all other $i \in \mathcal{I} \setminus \{j, k\}$, and furthermore, none of x_j, y_j, x_k, y_k is in the interval $[0, \theta]$. Consider the following axioms.

Restricted inequality neutrality. There is some $\theta > 0$ such that, for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, if $\mathbf{y}$ is a θ-restricted Pigou–Dalton transform of $\mathbf{x}$, then $\mathbf{y} \approx \mathbf{x}$.
Restricted inequality aversion. There is some $\theta > 0$ such that, for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, if $\mathbf{y}$ is a θ-restricted Pigou–Dalton transform of $\mathbf{x}$, then $\mathbf{y} \succeq \mathbf{x}$.
Restricted strict inequality aversion. There is some $\theta > 0$ such that, for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, if $\mathbf{y}$ is a θ-restricted Pigou–Dalton transform of $\mathbf{x}$, then $\mathbf{y} \succ \mathbf{x}$.

This might seem like a rather stingy version of inequality aversion, since it specifically excludes the wretched. But it allows us to avoid the Repugnant Conclusion.

Proposition 2.6 *Let $\succeq$ be a rank-additive possibilist SWO with the SWF* (2B).

(a) $\succeq$ *satisfies* Restricted inequality neutrality *if and only if there are linear functions $\phi^\pm : \mathbb{R}_\pm \longrightarrow \mathbb{R}_\pm$ and constants $\{c_n\}_{n=1}^\infty$ such that for all $n \in \mathbb{N}$, $\phi_n^- = \phi^-$ and $\phi_n^+(r) = \phi^+(r) + c_n$ for all $r \geqslant \theta$.*
(b) $\succeq$ *satisfies* Restricted inequality aversion *if and only if, for all $n, m \in \mathbb{N}$, all $r, s \in \mathbb{R}$ and all $\epsilon > 0$:*

- *If $r \geqslant \theta > 0 \geqslant s$, then $\phi_n^+(r + \epsilon) - \phi_n^+(r) \leqslant \phi_m^-(s) - \phi_m^-(s - \epsilon)$.*
- *If $n < m$ and $r > s > \epsilon + \theta > 0$, then $\phi_n^+(r + \epsilon) - \phi_n^+(r) \leqslant \phi_m^+(s) - \phi_m^+(s - \epsilon)$.*
- *If $n > m$ and $s < r < -\epsilon < 0$, then $\phi_n^-(r + \epsilon) - \phi_n^-(r) \leqslant \phi_m^-(s) - \phi_m^-(s - \epsilon)$.*

In particular, for all $r < 0$, the sequence $\{\phi_n^-(r)\}_{n=1}^\infty$ is nonincreasing.

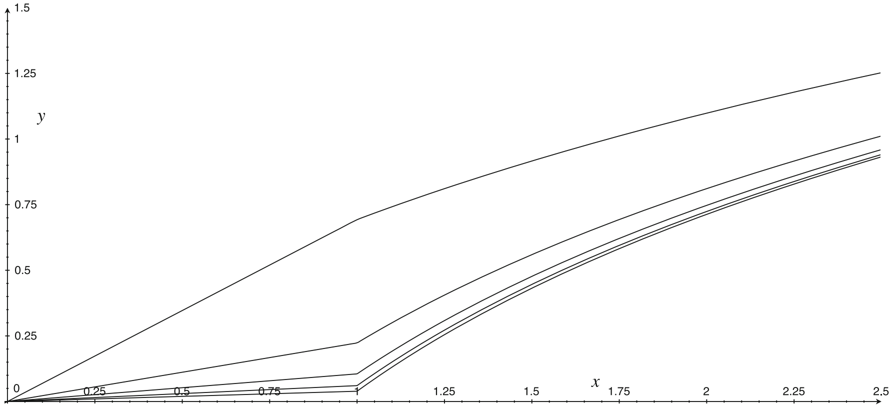

Fig. 2 Functions $\phi_n^+(r)$ in Example 2.7, for $n \in \{1, \ldots, 5\}$

(c) $\succeq$ *satisfies* Restricted strict inequality aversion *if and only if all the statements in part* (b) *hold with strict inequalities. In this case, for all* $r \in \mathbb{R}_-$, *the sequence* $\{\phi_n^-(r)\}_{n=1}^\infty$ *is strictly decreasing.*

Note that Proposition 2.6 does not require the sequence $\{\phi_n^+(r)\}_{n=1}^\infty$ to be nondecreasing for any $r > 0$. In effect, ϕ_n^+ must be inequality-averse for "sufficiently large" lifetime utilities (those above the threshold θ), but to block the Repugnant Conclusion, ϕ_n^+ must become increasingly inequality-*seeking* for "small" positive lifetime utilities (those in $[0, \theta]$), as $n \to \infty$. This is because the rank-additive SWF (2B) must assign rapidly decreasing marginal value to adding more wretched people to an already very populated world. But to respect Positive expansion, $\succeq$ must still regard these wretched new lives as a net improvement, as long as they are lives worth living. The only way to reconcile these two conflicting imperatives is for the slope of ϕ_n^+ near zero to decay to zero as $n \to \infty$.[13]

Example 2.7 For all $n \in \mathbb{N}$, let $a_n := \ln\left(1 + \frac{1}{n^2}\right)$, and then define

$$\phi_n^+(r) := \begin{cases} a_n\, r & \text{if} \quad r \in [0, 1]; \\ \ln\left(r + \frac{1}{n^2}\right) & \text{if} \quad r \geqslant 1. \end{cases}$$

(See Fig. 2.) If n is large, then $a_n \approx 1/n^2$ (by the Taylor expansion of $\ln(x)$ around $x = 1$). Thus, for all $r \in [0, 1]$, we have $\sum_{n=1}^\infty \phi_n^+(r) \approx r \cdot \sum_{n=1}^\infty \frac{1}{n^2}$, which is finite; thus, the hypothesis of Proposition 2.2(a) is satisfied, so the resulting RA SWO satisfies No Repugnant Conclusion (for any $r_0 < 1$). Meanwhile, for all $n \in \mathbb{N}$, we have

[13] Similarly, Roemer (2004) proposed an axiom he called *Triage*, which treats individuals differently depending on whether their utility is above or below a threshold corresponding to a "barely mediocre" life. But Roemer was not concerned with population ethics; rather, he was concerned with reconciling conflicting intuitions about distributional ethics which apply at different levels of utility.

$$(\phi_n^+)'(r) = \frac{1}{r + 1/n^2}, \quad \text{for all } r \in (1, \infty).$$

Thus, ϕ_n^+ is concave increasing on $[1, \infty)$, and furthermore, if $n < m$ and $r \geqslant s$, then $(\phi_n^+)'(r) < (\phi_m^+)'(s)$. Thus, the second condition in Proposition 2.6(c) is satisfied (with $\theta := 1$). Observe that $(\phi_n^+)'(r) < 1$ for all $n \in \mathbb{N}$ and $r \in \mathbb{R}_+$. Thus, if $\phi^- : \mathbb{R}_- \longrightarrow \mathbb{R}_-$ is any concave, increasing, differentiable function such that $\phi^-(0) = 0$ and $(\phi^-)'(0) \geqslant 1$, and we define $\phi_n^- := \phi^-$ for all $n \in \mathbb{N}$, then the other two conditions of Proposition 2.6(c) are also satisfied. Thus, the resulting SWO satisfies Restricted strict inequality aversion. □

Related literature In possibilist population ethics, a lifetime utility of zero plays a special role, because it is the utility assigned to non-existent people. This makes it possible to axiomatize SWOs that treat positive and negative utilities differently—a possibility first recognized by Blackorby and Donaldson (1982) and further developed by Zank (2007). Like the results in this section, Zank axiomatically characterizes rank-dependent SWOs that assign special significance to zero utility, and hence treat positive and negative utilities differently. However, he works with a fixed population and uses different axioms than the ones used here.

3 Actualist social welfare orders

3.1 Definition and examples

As in Sect. 2, let $\mathcal{I}$ be an infinite set, whose elements represent all the people who could ever possibly exist. Let $\mathbb{R}_* := \mathbb{R} \sqcup \{\nexists\}$, where $\nexists$ is a special symbol representing "non-existence". Let $\mathbb{R}_*^{\mathcal{I}}$ be the set of all $\mathcal{I}$-indexed profiles $\mathbf{r} = (r_i)_{i \in \mathcal{I}}$ of $\mathbb{R}_*$. For all $i \in \mathcal{I}$, if $r_i = \nexists$, then this means i does not exist. On the other hand, if $r_i \in \mathbb{R}$, then interpret r_i as the *lifetime utility* of individual i, with the same interpretation as in Sect. 2: if $r_i > 0$, then i's life is worth living, if $r_i < 0$, then i's life is *not* worth living, and if $r_i = 0$, then i's life is indifferent (for her) to non-existence.

Let $\mathcal{X}_\propto$ be the set of all elements of $\mathbb{R}_*^{\mathcal{I}}$ where only finitely many entries are not equal to $\nexists$. (Some of these non-$\nexists$ entries may be zero.) An element of $\mathcal{X}_\propto$ represents a complete specification of all the people who will ever exist (I assume this number to be finite), and the lifetime utilities each of them. I will refer to elements of $\mathcal{X}_\propto$ as *social outcomes*. An *actualist social welfare order* is a preference order on $\mathcal{X}_\propto$.[14]

If $\pi : \mathcal{I} \longrightarrow \mathcal{I}$ is any bijection, then define $\pi^* : \mathbb{R}_*^{\mathcal{I}} \longrightarrow \mathbb{R}_*^{\mathcal{I}}$ by setting $\pi^*(\mathbf{r}) := (r_{\pi(i)})_{i \in \mathcal{I}}$ for all $\mathbf{r} = (r_i)_{i \in \mathcal{I}}$ in $\mathbb{R}_*^{\mathcal{I}}$. Clearly, $\pi(\mathcal{X}_\propto) = \mathcal{X}_\propto$, and π restricted to $\mathcal{X}_\propto$ defines a bijection from $\mathcal{X}_\propto$ to itself. We will be interested in SWOs satisfying the following axiom:

Anonymity. If $\pi : \mathcal{I} \longrightarrow \mathcal{I}$ is any bijection, and $\mathbf{x} \in \mathcal{X}_\propto$, then $\mathbf{x} \approx \pi^*(\mathbf{x})$.

[14] There is a risk of terminological confusion here: "moral actualism" has also been used to refer to the philosophical claim that ethical judgements should be based only on the interests of the people who actually exist. See Hare (2007) for a refutation of this position. This is not what I mean by the term.

For any $\mathbf{x} \in \mathcal{X}_\propto$, let $|\mathbf{x}|$ be the number of non-$\nexists$ entries in $\mathbf{x}$. In particular, let $\emptyset$ be the *empty world*: the unique element of $\mathcal{X}_\propto$ such that *all* entries are $\nexists$; then, $|\emptyset| = 0$. If $|\mathbf{x}| > 0$, then we say $\mathbf{x}$ is *nonempty*. For any $N \in \mathbb{N}$, let $\mathcal{X}_N := \{\mathbf{x} \in \mathbb{R}_*^{\mathcal{I}};\ |\mathbf{x}| = N\}$, and let $\mathbb{R}^{N\uparrow} := \{\mathbf{r} \in \mathbb{R}^N;\ r_1 \leqslant r_2 \leqslant \cdots \leqslant r_N\}$ be the set of all nondecreasing elements of $\mathbb{R}^N$. For any $\mathbf{x} \in \mathcal{X}_N$, let $\mathbf{x}^\uparrow := (x_1^\uparrow, x_2^\uparrow, \ldots, x_N^\uparrow) \in \mathbb{R}^{N\uparrow}$ be the N-dimensional vector consisting of all non-$\nexists$ entries of $\mathbf{x}$, listed in nondecreasing order. Let $\mathbb{R}^{\propto\uparrow} := \bigcup_{N=1}^{\infty} \mathbb{R}^{N\uparrow}$, and let $\succeq_*$ be a preference order on $\mathbb{R}^{\propto\uparrow}$. Then we can define an SWO $\succeq$ on $\mathcal{X}_\propto$ by the formula:

$$\text{for all } \mathbf{x}, \mathbf{y} \in \mathcal{X}_\propto \qquad \big(\mathbf{x} \succeq \mathbf{y}\big) \iff \big(\mathbf{x}^\uparrow \succeq_* \mathbf{y}^\uparrow\big). \tag{3A}$$

It is easy to see that $\succeq$ satisfies Anonymity: if $\pi : \mathcal{I} \longrightarrow \mathcal{I}$ is any bijection, and $\mathbf{y} = \pi^*(\mathbf{x})$, then $\mathbf{y}^\uparrow = \mathbf{x}^\uparrow$, so that $\mathbf{x} \approx \mathbf{y}$. Conversely, if $\succeq$ is an SWO on $\mathcal{X}_\propto$ satisfying Anonymity, then there is a unique preference order $\succeq_*$ on $\mathbb{R}^{\propto\uparrow}$ satisfying formula (3A). In other words, there is a natural bijective correspondence between preference orders on $\mathbb{R}^{\propto\uparrow}$ and SWOs on $\mathcal{X}_\propto$ satisfying Anonymity. Thus, we can work directly with preference orders on $\mathbb{R}^{\propto\uparrow}$.

For all $n \in \mathbb{N}$, let $\phi_n : \mathbb{R} \longrightarrow \mathbb{R}$ be a continuous, increasing function. Consider the social welfare function $W : \mathcal{X}_\propto \longrightarrow \mathbb{R}$ defined as follows:

$$W(\emptyset) := 0, \quad \text{and} \quad W(\mathbf{x}) := \sum_{n=1}^{|\mathbf{x}|} \phi_n(x_n^\uparrow), \ \text{ for all nonempty } \mathbf{x} \in \mathcal{X}_\propto. \tag{3B}$$

This is called an *ascending rank-additive* (ARA) social welfare function. The social welfare order it represents is an *ascending rank-additive* SWO. For example:

- Suppose $c \in \mathbb{R}$, and $\phi_n(r) = r - c$ for all $n \in \mathbb{N}$ and all $r \in \mathbb{R}$. Then formula (3B) yields the *critical-level utilitarian* SWF. In particular, if $c = 0$, then we get the classical utilitarian SWF. If $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ is a continuous, increasing function, and $\phi_n = \phi$ for all $n \in \mathbb{N}$, then (3B) yields a *generalized utilitarian* SWF:

$$W(\emptyset) := 0, \quad \text{and} \quad W(\mathbf{x}) := \sum_{n=1}^{|\mathbf{x}|} \phi(x_n^\uparrow), \ \text{ for all nonempty } \mathbf{x} \in \mathcal{X}_\propto. \tag{3C}$$

- Let $\{a_n\}_{n=1}^{\infty}$ be a decreasing sequence of positive constants, and suppose $\phi_n(r) = a_n\, r$ for all $n \in \mathbb{N}$ and all $r \in \mathbb{R}$. Then formula (3B) yields an *ascending rank-weighted utilitarian* SWF.
- More generally, let $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ be a continuous, increasing function, and suppose $\phi_n(r) = a_n\, \phi(r)$ for all $n \in \mathbb{N}$ and all $r \in \mathbb{R}$. Then formula (3B) yields an *ascending rank-weighted generalized utilitarian* SWF:

$$W(\mathbf{x}) := \sum_{n=1}^{|\mathbf{x}|} a_n\, \phi(x_n^\uparrow), \quad \text{for all } \mathbf{x} \in \mathcal{X}_\propto. \tag{3D}$$

These have been studied by Asheim and Zuber (2017). In particular, let $\beta \in (0, 1)$, and for all $n \in \mathbb{N}$, let $\phi_n := \beta^n\, \phi$. Then formula (3D) becomes a *rank-discounted* *generalized utilitarian* SWF, which was axiomatically characterized by Asheim and Zuber (2014):

$$W(\mathbf{x}) := \sum_{n=1}^{|\mathbf{x}|} \beta^n \, \phi(x_n^{\uparrow}), \quad \text{for all } \mathbf{x} \in \mathcal{X}_{\propto}. \tag{3E}$$

In these examples, I do not assume that $\phi_n(0) = 0$. In other words, I do *not* assume that the existence of an individual with a neutral level of lifetime utility is ethically equivalent to her non-existence. In fact, even if $\phi_n(0) = 0$ for all $n \in \mathbb{N}$, this would not be the case: introducing a new person with zero lifetime utility can change the rankings of people who already exist, thereby changing overall social welfare in a complex way. Thus, ARA SWOs are fundamentally different from the possibilist rank-additive SWOs introduced in Sect. 2; they typically do *not* satisfy either Positive expansion or Negative expansion.

However, as noted by Asheim and Zuber (2014, 2016, 2017), ARA SWOs are attractive because they can avoid the Repugnant Conclusion while exhibiting inequality aversion at all welfare levels. To see this, consider the ascending rank-weighted generalized utilitarian SWF (3D). For simplicity, suppose $\phi(r) = r$ for all $r \in \mathbb{R}$. If the sequence $\{a_n\}_{n=1}^{\infty}$ is decreasing, then this SWF is inequality-averse, because it assigns lower marginal social welfare to the lifetime utilities of more fortunate individuals (who appear higher in the ranking). Furthermore, Asheim and Zuber (2017, Proposition 6) show that this SWF avoids the Repugnant Conclusion if and only if $\sum_{n=1}^{\infty} a_n < \infty$. Clearly, this summability condition is compatible with $\{a_n\}_{n=1}^{\infty}$ being a decreasing sequence—for example, it is satisfied by the rank-discounted generalized utilitarian SWF (3E). I generalize this result in Proposition 3.2.

3.2 Axiomatic characterization

I will characterize ARA SWOs with six axioms. The first one is Anonymity. The next three are quite standard and also appeared in Sect. 2.2. To state these axioms, suppose an SWO $\succeq$ on $\mathcal{X}_{\propto}$ satisfies Anonymity. Then it can represented by a preference order $\succeq_*$ on $\mathbb{R}^{\propto\uparrow}$. For any $N \in \mathbb{N}$, recall that $\mathbb{R}^{N\uparrow} \subset \mathbb{R}^{\propto\uparrow}$; let $\succeq_N$ be the restriction of $\succeq_*$ to $\mathbb{R}^{N\uparrow}$. Note that $\mathbb{R}^{N\uparrow}$ is a closed, convex subset of $\mathbb{R}^N$; endow it with the subspace topology it inherits from $\mathbb{R}^N$. The next two axioms concern the preference orders $\{\succeq_N\}_{N=1}^{\infty}$.

Continuity. For every $\mathbf{x} \in \mathcal{X}_{\propto}$, and every $N \in \mathbb{N}$, the upper contour sets $\{\mathbf{y}^{\uparrow};\ \mathbf{y} \in \mathcal{X}_N \text{ and } \mathbf{x} \preceq \mathbf{y}\}$ and the lower contour sets $\{\mathbf{y}^{\uparrow};\ \mathbf{y} \in \mathcal{X}_N \text{ and } \mathbf{x} \succeq \mathbf{y}\}$ are closed subsets of $\mathbb{R}^{N\uparrow}$.

Pareto. For every $N \in \mathbb{N}$ and for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{N\uparrow}$, if $x_n \geqslant y_n$ for all $n \in [1 \ldots N]$, then $\mathbf{x} \succeq_N \mathbf{y}$. If, furthermore, $x_n > y_n$ for some $n \in [1 \ldots N]$, then $\mathbf{x} \succ_N \mathbf{y}$.

These axioms have the same justification as the corresponding axioms in Sect. 2. Note that Continuity is slightly stronger than requiring the orders $\succeq_N$ to be continuous: it also requires closure of contour sets determined by elements outside of $\mathcal{X}_N$.[15]

[15] Blackorby et al. (2001, 2005) and Asheim and Zuber (2017) refer to a similar axiom as *Extended continuity*.

To formulate the last axiom, we need some notation. Let $\mathbf{x} \in \mathcal{X}_\propto$, let $N := |\mathbf{x}|$, let $n \in [1 \ldots N]$, and let $b \in \mathbb{R}$ be such that $x^\uparrow_{n-1} \leqslant b \leqslant x^\uparrow_{n+1}$.[16] Let $b_{(n)}\mathbf{x}$ be the unique element $\mathbf{y} \in \mathcal{X}$ such that $|\mathbf{y}| = |\mathbf{x}|$, $y^\uparrow_n = b$, and $y^\uparrow_m = x^\uparrow_m$ for all other $m \in [1 \ldots N] \setminus \{n\}$.[17]

Now let $n < m \in \mathbb{N}$, and let $a < b < c < d \in \mathbb{R}$. I will write $(a \overset{n}{\rightsquigarrow} b) \cong (c \overset{m}{\rightsquigarrow} d)$ if there exists $\mathbf{x} \in \mathcal{X}_\propto$ such that $x^\uparrow_n = a$, $x^\uparrow_m = c$, $b_{(n)}\mathbf{x}$, and $d_{(m)}\mathbf{x}$ are well defined, and $b_{(n)}\mathbf{x} \approx d_{(m)}\mathbf{x}$. This means that switching a to b in coordinate n is "ethically equivalent" to switching c to d in coordinate m. If $\succeq$ is represented by a SWF W, then $(a \overset{n}{\rightsquigarrow} b) \cong (c \overset{m}{\rightsquigarrow} d)$ if the change in W induced by switching a to b in the nth coordinate is exactly equal to the change in W induced by switching c to d in the mth coordinate. If W has an ascending rank-additive representation (3B), then $(a \overset{n}{\rightsquigarrow} b) \cong (c \overset{m}{\rightsquigarrow} d)$ if and only if $\phi_n(b) - \phi_n(a) = \phi_m(d) - \phi_m(c)$. Here is the last axiom:

> Trade-off Consistency. For any $n < m \in \mathbb{N}$, and any $a < b < c < d \in \mathbb{R}$ such that $(a \overset{n}{\rightsquigarrow} b) \cong (c \overset{m}{\rightsquigarrow} d)$, and any $\mathbf{y}, \mathbf{z} \in \mathcal{X}_\propto$, such that $y^\uparrow_n = a$, $y^\uparrow_{n-1} \leqslant b \leqslant y^\uparrow_{n+1}$, $z^\uparrow_m = c$, and $z^\uparrow_{m-1} \leqslant d \leqslant z^\uparrow_{m+1}$, if $\mathbf{y} \approx \mathbf{z}$, then $b_{(n)}\mathbf{y} \approx d_{(m)}\mathbf{z}$.

Note that this axiom does *not* assume that $|\mathbf{y}| = |\mathbf{z}|$. It says: if the act of switching a to b in coordinate n is "ethically equivalent" to the act of switching c to d in coordinate m when both switches are applied to the same outcome $\mathbf{x}$, then this same ethical equivalence should also be observed when these switches are applied to two *different* outcomes $\mathbf{y}$ and $\mathbf{z}$, possibly with different population sizes. Finally, we need the following structural condition.

> Neutral population growth. For all $N \in \mathbb{N}$, there exists some $\mathbf{x} \in \mathcal{X}_N$ such that $\mathbf{x} \approx \emptyset$.

This condition is natural and easily satisfied. For example, if $\succeq$ is a rank-weighted generalized utilitarian SWO as in (3D), then it satisfies Neutral population growth if and only if $\phi(r) = 0$ for some $r \in \mathbb{R}$. Meanwhile, if $\succeq$ is an ARA SWO represented by (3B), then it satisfies Neutral population growth if ϕ_1 is unbounded below, while ϕ_n takes at least some positive values for all $n \geqslant 2$. Here is the second main result of the paper.

Theorem 2 *Let $\succeq$ be an actualist SWO satisfying* Neutral population growth *on $\mathcal{X}_\propto$. Then $\succeq$ satisfies* Anonymity*,* Continuity*,* Pareto*, and* Trade-off Consistency *if and only if it is ascending rank additive. In the representation* (3B)*, the functions $\{\phi_n\}_{n=1}^\infty$ are unique up to multiplication by a common scalar.*

Proof sketch For all $N \in \mathbb{N}$, use Trade-off Consistency to show that $\succeq_N$ satisfies a limited separability axiom called *Coordinate Independence*[18] in a neighbourhood around each point in $\mathbb{R}^{N}_{\uparrow}$. Then use a result of Wakker (1988) to obtain "local" additive representations in a neighbourhood of each point in $\mathbb{R}^{N}_{\uparrow}$. Then, as in the proof of

[16] Here we adopt the notational convention that $x^\uparrow_0 := -\infty$ and $x^\uparrow_{N+1} = \infty$.

[17] Note that "$b_{(n)}\mathbf{x}$" is not well defined unless $x^\uparrow_{n-1} \leqslant b \leqslant x^\uparrow_{n+1}$.

[18] This is like Separability in Sect. 2.2, but in the case when $\mathcal{J}_\pm$ is the complement of a single coordinate.

Theorem 1, use a result of Chateauneuf and Wakker (1993) to get a global additive representation on all of $\mathbb{R}^{N\uparrow}$. Next, use Trade-off Consistency to show that, for any $N < M$, the N utility functions in the additive representation on $\mathbb{R}^{N\uparrow}$ are actually the first N utility functions for the additive representation on $\mathbb{R}^{M\uparrow}$. Finally, use Trade-off Consistency again to show that these additive representations also correctly account for comparisons between outcomes with different population sizes. For details, see Appendix B.

Independence of the axioms One can demonstrate the independence of Anonymity, Continuity, Pareto using examples very similar to those at the end of Sect. 2.2. Meanwhile, one can show the independence of Trade-off Consistency using an example very similar to the example for Separability from Sect. 2.2.

3.3 Further results

Critical levels Let $\mathbf{x} \in \mathcal{X}_\propto$ and let $c \in \mathbb{R}$. In the terminology of Blackorby et al. (2005), c is a *critical level* for $\mathbf{x}$ if adding a new person with lifetime utility c to $\mathbf{x}$ is an ethically neutral act. By the Pareto axiom, such a critical level is unique, if it exists. For example, in the classical utilitarian SWO, $c = 0$ for all $\mathbf{x} \in \mathcal{X}_\propto$. In the average utilitarian SWO, c is the average lifetime utility in $\mathbf{x}$. The ARA SWOs characterized in Theorem 2 do not necessarily possess such critical levels for every social outcome. In other words, they do not necessarily satisfy the following axiom:

Critical levels. For any $\mathbf{x} \in \mathcal{X}_\propto$, there exists $c \in \mathbb{R}$ (depending on $\mathbf{x}$) with $\mathbf{x} \approx \mathbf{x} \uplus c$.

This axiom says that there is no outcome $\mathbf{x}$ so bad that adding *any* new person to $\mathbf{x}$ is always considered an improvement, or so good that adding *any* new person to $\mathbf{x}$ is always considered a deterioration. Suppose $\succeq$ is an ARA SWO defined by a collection of functions $\boldsymbol{\phi} := \{\phi_n\}_{n=1}^{\infty}$. To ensure that $\succeq$ satisfies Critical levels, we must impose some conditions on $\boldsymbol{\phi}$. One might think that it is sufficient to require, for all $n \in \mathbb{N}$, the existence of some $c_n \in \mathbb{N}$ with $\phi_n(c_n) = 0$. But this is not quite sufficient, as we will now see. For all $n \in \mathbb{N}$, define the function $\delta\phi_n : \mathbb{R} \longrightarrow \mathbb{R}$ by $\delta\phi_n(r) := \phi_{n+1}(r) - \phi_n(r)$. Then define

$$S(\boldsymbol{\phi}) := \sup \left\{ \sum_{n=1}^{N} \delta\phi_n(x_n) \; ; \; N \in \mathbb{N} \text{ and } x_1 \leqslant x_2 \leqslant \cdots \leqslant x_N \right\}. \tag{3F}$$

Proposition 3.1 *Let $\succeq$ be an ARA SWO on $\mathcal{X}_\propto$ with representation (3B), such that for all $n \in \mathbb{N}$, there is some $c_n \in \mathbb{N}$ with $\phi_n(c_n) = 0$. The following statements are equivalent:*

(a) $\succeq$ *satisfies* Critical levels.
(b) $\inf(\phi_1(\mathbb{R})) \leqslant -S(\boldsymbol{\phi})$, *and if* $\inf(\phi_1(\mathbb{R})) = -S(\boldsymbol{\phi})$, *then the supremum in formula* (3F) *is never obtained.*

In most cases, the condition in Proposition 3.1 is easily satisfied. For example, if ϕ_1 is unbounded below (so that $\inf(\phi_1(\mathbb{R})) = -\infty$), then the condition is automatically true. Meanwhile, in a generalized utilitarian SWO (3C), we have $S(\boldsymbol{\phi}) = 0$, so the condition simply requires that that $\inf(\phi_1(\mathbb{R})) < 0$.

Suppose that $\succeq$ is as in Proposition 3.1. Then for any $N \in \mathbb{N}$ and $\mathbf{x} \in \mathcal{X}_\propto$, if $|\mathbf{x}| = N - 1$ and $\max(\mathbf{x}) \leqslant c_N$, then $\mathbf{x} \approx \mathbf{x} \uplus c_N$. In other words, adding a person with lifetime utility c_N to the world is an ethically neutral act, as long as everyone who already exists has an even lower level of lifetime utility. This is similar to the axiom *Existence of a critical level* employed by Asheim and Zuber (2014) in their axiomatic characterization of rank-discounted generalized utilitarian SWOs, but weaker: Asheim and Zuber additionally require that $c_N = c_M$ for all $N, M \in \mathbb{N}$.

Inequality and the Repugnant Conclusion As Asheim and Zuber (2014) noted, an ARA SWO can reconcile inequality aversion with evasion of the Repugnant Conclusion by assigning lower marginal social welfare to the lifetime utility of the better-off individuals in any social outcome. The next result makes this precise. It parallels Propositions 2.2 and 2.4.

Proposition 3.2 *Let $\succeq$ be an ARA SWO with the SWF* (3B).

(a) $\succeq$ *satisfies* No Repugnant Conclusion *if and only if there exists $r > 0$ such that* $\displaystyle\sum_{n=1}^{\infty} \phi_n(r) < \infty$.

(b) $\succeq$ *satisfies* Inequality aversion *if and only if for all $n, m \in \mathbb{N}$ with $n \geqslant m$, all $r, s \in \mathbb{R}$ with $r \geqslant s$, and all $\epsilon > 0$, we have $\phi_n(r+\epsilon) - \phi_n(r) \leqslant \phi_m(s) - \phi_m(s-\epsilon)$. In particular, for all $r \in \mathbb{R}_+$, we have $\phi_1(r) \geqslant \phi_2(r) \geqslant \phi_3(r) \geqslant \cdots$ Furthermore, if $\{\phi_n\}_{n=1}^{\infty}$ are differentiable, then for any nondecreasing sequence $r_1 \leqslant r_2 \leqslant r_3 \leqslant \cdots$ of real numbers, we have $\phi_1'(r_1) \geqslant \phi_2'(r_2) \geqslant \phi_3'(r_3) \geqslant \cdots$*

(c) $\succeq$ *satisfies* Strict inequality aversion *if and only if all the statements in part* (b) *hold with strict inequalities.*

There are also versions of Proposition 2.2(b) and 2.3 for ARA SWO (for avoiding utility monsters and the St. Petersburg Paradox), but they are obvious, and are left to the reader.

Example 3.3 Let $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ be a concave increasing function, let $\{a_n\}_{n=1}^{\infty}$ be a nonincreasing sequence of positive constants, and suppose W is the ascending rank-weighted generalized utilitarian SWF (3D). Then Proposition 3.2(b) says that $\succeq$ satisfies Inequality aversion. If $\sum_{n=1}^{\infty} a_n < \infty$, then Asheim and Zuber (2017) say that $\succeq$ is *proper*. In this case, Proposition 3.2(a) says that $\succeq$ satisfies No Repugnant Conclusion. In particular, if $\beta \in (0, 1)$, and W is the rank-discounted generalized utilitarian SWF (3E), then $\succeq$ satisfies both Strict inequality aversion and No Repugnant Conclusion. □

Tyrannies of aggregation and nonaggregation For any $\mathbf{x} \in \mathcal{X}_\propto$, let $\mathcal{I}(\mathbf{x}) := \{i \in \mathcal{I};\ x_i \neq \nexists\}$ be the set of people who exist in the outcome $\mathbf{x}$. Fleurbaey and Tungodden (2010) have proposed the following axiom:

> Minimal Aggregation. For some $N \in \mathbb{N}$, and all $\mathbf{x} \in \mathcal{X}_N$, and all $i \in \mathcal{I}(\mathbf{x})$, there exists some $\alpha > \beta > 0$ such that, for any other $\mathbf{y} \in \mathcal{X}_\propto$ with $\mathcal{I}(\mathbf{y}) = \mathcal{I}(\mathbf{x})$, if $x_i \geqslant y_i \geqslant x_i - \beta$, while $y_j \geqslant x_j + \alpha$ for all other $j \in \mathcal{I}(\mathbf{x})$, then $\mathbf{y} \succeq \mathbf{x}$.

This says that there is at least one situation where it is considered acceptable for one person (namely i) to make a small sacrifice (at most β) so that everyone else can gain a larger amount (at least α). This rules out the maximin and leximin SWFs, which give *absolute* priority to the worst-off. In the terminology of Fleurbaey and Tungodden, it excludes the *tyranny of nonaggregation*.

For any $\mathbf{x} \in \mathcal{X}_\propto$, let $\underline{\mathcal{I}}(\mathbf{x}) := \{i \in \mathcal{I}(\mathbf{x});\ \ x_i \leqslant x_j$ for all $j \in \mathcal{I}(\mathbf{x})\}$; these are the worst-off people in the outcome $\mathbf{x}$. Let $\overline{\mathcal{I}}(\mathbf{x}) := \{i \in \mathcal{I}(\mathbf{x});\ \ x_i \geqslant x_j$ for all $j \in \mathcal{I}(\mathbf{x})\}$; these are the best-off people in $\mathbf{x}$. Fleurbaey and Tungodden (2010) also propose the next axiom:

> Mild Nonaggregation. For all $r, q \in \mathbb{R}$ with $r > q$, there exists some $\alpha > \beta > 0$ such that, for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}_\propto$ with $\mathcal{I}(\mathbf{x}) = \mathcal{I}(\mathbf{y})$, if there exists $i \in \underline{\mathcal{I}}(\mathbf{y})$ with $y_i \leqslant q$ and $x_i \geqslant y_i + \alpha$, while for all other $j \in \mathcal{I}(\mathbf{x})$, either $x_j = y_j$ or $j \in \overline{\mathcal{I}}(\mathbf{x}) \cap \overline{\mathcal{I}}(\mathbf{y})$, $y_j \geqslant r$, and $x_j \geqslant y_j - \beta$, then $\mathbf{x} \succeq \mathbf{y}$.

This is an egalitarian principle which says that, if one of the worst-off people (namely i) has the opportunity to gain a large enough amount (at least α), then it is acceptable to require all the best-off people to make a small sacrifice (at most β). This rules out principles such as utilitarianism, in which a tiny welfare gain for every member of a large group (say, a billion people) could in principle justify a huge sacrifice (say, a painful death) for one person. In the terminology of Fleurbaey and Tungodden, it excludes the *tyranny of aggregation*. Note that the axiom allows us to make α arbitrarily large and β arbitrarily small, depending on the values of q and r; in that sense it is a fairly mild principle. However, α and β do *not* depend on the size of $\mathcal{I}(\mathbf{x})$; thus, even if $\overline{\mathcal{I}}(\mathbf{x}) \cap \overline{\mathcal{I}}(\mathbf{y})$ has a quadrillion people, their β-sized gains are insufficient to offset i's α-sized loss.

For any $\mathbf{x} \in \mathcal{X}_\propto$ and $N \in \mathbb{N}$, let $\mathbf{x}^N$ denote a social outcome with $|\mathbf{x}^N| = N\,|\mathbf{x}|$, such that every person in $\mathbf{x}$ is "replicated" N times in $\mathbf{x}^N$. (If an SWO satisfies Anonymity, it does not matter how we arrange these N replicated copies). An SWO $\succeq$ is *replication-invariant* if, for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}_\propto$ with $\mathbf{x} \succeq \mathbf{y}$, we have $\mathbf{x}^N \succeq \mathbf{y}^N$ for all $N \in \mathbb{N}$. It is easily seen that rank-additive SWOs are *not* replication-invariant, in general.

Fleurbaey and Tungodden prove that Minimal Aggregation and Mild Nonaggregation are incompatible in any replication-invariant SWO.[19] But they note that these axioms *are* compatible if we give up replication invariance; in particular, they are both satisfied by rank-discounted utilitarian SWFs like (3E). The next result makes this precise.

Proposition 3.4 (a) *Any ARA SWO satisfies* Minimal Aggregation.

(b) *Let $\succeq$ be an ARA SWO with the SWF (3B). Suppose $\{\phi_n\}_{n=1}^\infty$ are concave and differentiable, and for all $r \in \mathbb{R}$, there exists $A \in \mathbb{R}_+$ such that $\sum_{n=N+1}^\infty \phi'_n(r) < A \cdot \phi'_m(r)$, for any $m, N \in \mathbb{N}$ with $m \leqslant N$. Then $\succeq$ satisfies* Mild Nonaggregation.

The condition in Proposition 3.4(b) implies that the sequence $\{\phi'_n(r)\}_{n=1}^\infty$ decays quickly enough that $\sum_{n=1}^\infty \phi'_n(r) < \infty$. But it says more: it says that this sequence

[19] Fleurbaey and Tungodden are not concerned with population ethics, so they only consider social welfare orders that compare social outcomes having the *same* population. I have reformulated their axioms in the notation of this paper.

decays with roughly "exponential" speed. For example, suppose there is some $b \in (0,1)$ such that $\phi'_{n+1}(r) \leqslant b \cdot \phi'_n(r)$ for all $n \in \mathbb{N}$ (e.g. $\phi'_n(r) = 1/b^n$ for all $n \in \mathbb{N}$); then, the condition in Proposition 3.4(b) is satisfied, with $A := b/(1-b)$. Asheim and Zuber (2017, Proposition 5) have proved a result similar to Proposition 3.4 for ascending rank-weighted generalized utilitarian SWFs of type (3D).

Top-independence and bottom-independence For any $\mathbf{x} \in \mathcal{X}_\propto$, let $\max(\mathbf{x})$ be the maximal lifetime utility of any person in the social outcome $\mathbf{x}$. (Equivalently, if $\mathbf{x}^\uparrow = (x_1^\uparrow, \ldots, x_N^\uparrow)$, then $\max(\mathbf{x}) = x_N^\uparrow$.) All ARA SWOs clearly satisfy the next axiom.

> Top-independence. For all $\mathbf{x}, \mathbf{y} \in \mathcal{X}_\propto$ with $|\mathbf{x}| = |\mathbf{y}|$ and all $z \in \mathbb{R}$ with $z \geqslant \max\{\max(\mathbf{x}), \max(\mathbf{y})\}$, we have $\mathbf{x} \succeq \mathbf{y}$ if and only if $\mathbf{x} \uplus z \succeq \mathbf{y} \uplus z$.

This axiom has a similar justification to Separability or the axiom of Top-independence in good worlds from Sect. 2.3: it says that, when comparing two social outcomes, we can disregard the utility—or even the *existence*—of someone who is indifferent between these outcomes, as long as she would be the happiest person in both outcomes. Unfortunately, ARA SWOs do *not*, in general, satisfy the analogous property involving unhappy people:

> Bottom-independence. For all $\mathbf{x}, \mathbf{y} \in \mathcal{X}_\propto$ with $|\mathbf{x}| = |\mathbf{y}|$ and all $z \in \mathbb{R}$ with $z \leqslant \min\{\min(\mathbf{x}), \min(\mathbf{y})\}$, we have $\mathbf{x} \succeq \mathbf{y}$ if and only if $\mathbf{x} \uplus z \succeq \mathbf{y} \uplus z$.

One way to obtain SWFs satisfying Bottom-independence is to apply a formula like (3B) to *decreasing* rather than increasing sequences. Formally, for any $N \in \mathbb{N}$, let $\mathbb{R}^{N\downarrow} := \{\mathbf{r} \in \mathbb{R}^N;\ r_1 \geqslant \cdots \geqslant r_N\}$ be the set of all nonincreasing elements of $\mathbb{R}^N$. Let $\mathbb{R}^{\propto\downarrow} := \bigcup_{N=1}^{\infty} \mathbb{R}^{N\downarrow}$. For any $\mathbf{x} \in \mathcal{X}_N$, let $\mathbf{x}^\downarrow := (x_1^\downarrow, \ldots, x_N^\downarrow) \in \mathbb{R}^{N\downarrow}$ be the N-dimensional vector of all non-$\nexists$ entries of $\mathbf{x}$, listed in nonincreasing order. For all $n \in \mathbb{N}$, let $\phi_n : \mathbb{R} \longrightarrow \mathbb{R}$ be continuous and increasing. Define $W : \mathcal{X}_\propto \longrightarrow \mathbb{R}$ by:

$$W(\emptyset) \ := \ 0, \quad \text{and} \quad W(\mathbf{x}) \ := \ \sum_{n=1}^{|\mathbf{x}|} \phi_n(x_n^\downarrow), \quad \text{for all nonempty } \mathbf{x} \in \mathcal{X}_\propto.$$

This is called a *descending rank-additive* (DRA) social welfare function. These are axiomatically characterized by a result very similar to Theorem 2, except that the axioms Pareto, Continuity, and Trade-off Consistency are applied to orderings defined on $\mathbb{R}^{N\downarrow}$ rather than $\mathbb{R}^{N\uparrow}$ (for all $N \in \mathbb{N}$). However, DRA SWOs are less appealing than ARA SWOs. As observed in Proposition 3.2, ARA SWOs can simultaneously Strict inequality aversion and No Repugnant Conclusion. But DRA SWOs cannot. Indeed, for a DRA SWO to satisfy No Repugnant Conclusion, it must be inequality-*promoting*, which is much less attractive. Fortunately, there is another way to obtain Bottom-independence, while preserving the other desirable properties of ARA SWOs.

Corollary 3.5 *Let $\succeq$ be an actualist SWO on $\mathcal{X}_\propto$. Then $\succeq$ satisfies the axioms of Theorem 2 and also* Bottom-independence *if and only if it is rank-discounted generalized* *utilitarian, as in formula* (3E). *In this representation, β is unique and ϕ is unique up to multiplication by a positive constant.*

This is similar to the main result of Asheim and Zuber (2014), except that they do not require Trade-off Consistency, but instead employ versions of Top-independence and Bottom-independence,[20] an axiom positing the existence of egalitarian equivalents, and a slightly stronger form of Critical levels.

Related literature Ebert (1988) axiomatizes a class of ascending rank-additive SWOs, but he works with populations of fixed size, and uses different axioms. Bleichrodt et al. (2008) obtain a very general axiomatization of rank-additive utility representations. Although their intended application is descriptive decision theory, they note that their result can also be applied to social welfare. They also use continuity, monotonicity (i.e. Pareto), a richness condition analogous to Neutral population growth, and a "Trade-off Consistency" axiom, but their version of this axiom is different than the one in this paper.

3.4 Actualism versus possibilism

Proposition 3.1 shows that ARAs do not generally satisfy Critical levels. Even when they do, the critical level depends on the social outcome under consideration, and is not typically zero. But suppose that $\succeq$ was an actualist SWO where the critical level was zero for *every* social outcome. Thus, for any $\mathbf{x} \in \mathcal{X}_\propto$, if $\mathbf{x}'$ is obtained from $\mathbf{x}$ by converting any finite number of $\nexists$ components to zeros, then $\mathbf{x} \approx \mathbf{x}'$. Let $\mathcal{X}$ be the space of "possibilist" social outcomes defined in Sect. 2.1, and let $\Phi(\mathbf{x})$ be the element of $\mathcal{X}$ obtained by converting *all* the $\nexists$ components in $\mathbf{x}$ to zeros. This defines a surjection $\Phi : \mathcal{X}_\propto \longrightarrow \mathcal{X}$. We can then define a possibilist SWO $\succeq'$ on $\mathcal{X}$ as follows: for any $\mathbf{x}', \mathbf{y}' \in \mathcal{X}$, stipulate that $\mathbf{x}' \succeq' \mathbf{y}'$ if and only if $\mathbf{x} \succeq \mathbf{y}$ for some $\mathbf{x}, \mathbf{y} \in \mathcal{X}_\propto$ such that $\Phi(\mathbf{x}) = \mathbf{x}'$ and $\Phi(\mathbf{y}) = \mathbf{y}'$. (This is well defined independent of the choice of $\mathbf{x}$ and $\mathbf{y}$, precisely because the critical level of $\succeq$ is zero for all outcomes.)

Given this construction, it seems that possibilist SWOs are just a special case of actualist SWOs—namely those with a universal zero critical level. Thus, it seems redundant to introduce the separate framework of Sect. 2. It is also puzzling that the axioms Anonymity, Pareto, Continuity, and Separability were enough to axiomatically characterize rank-additive SWOs in the possibilist framework (Theorem 1), when these four axioms are *not* enough in the actualist framework. (The axiom of Trade-off Consistency is stronger than Separability.) Furthermore, the class of ARA SWOs we eventually *do* obtain in Theorem 2 is substantially different in nature.

However, a closer inspection reveals that the possibilist framework is not "just" a special case of the actualist framework. Its additional structure allows for logically weaker versions of key axioms and leads to different conclusions. For example, the possibilist Continuity axiom is *not* just the actualist Continuity axiom "projected" through the mapping Φ—it is a weaker axiom. To see this, note that the possibilist Continuity axiom makes a rigid distinction between the positive and negative compo-

[20] Asheim and Zuber's versions of Top- and Bottom-independence are called *Existence independence of the best off* and *Existence independence of the worst off*; they allow $\mathbf{x}$ and $\mathbf{y}$ to have different populations.

nents of a social outcome; it essentially says that the SWO is continuous with respect to small perturbations of individual utility levels, *as long as* these perturbations do not switch a positive utility to a negative one, or vice versa. The actualist Continuity axiom makes no such distinction. For similar reasons, the possibilist Separability axiom is weaker than the axiom that would be obtained by formulating an analogous Separability axiom in the actualist framework, and then "projecting" this axiom through Φ. There is a further difference between the two Continuity axioms (already noted in Sect. 3.2). Both axioms require closed contour sets in the set $\mathcal{X}_N$ of outcomes of size N (for all $N \in \mathbb{N}$). But the possibilist Continuity axiom only considers contour sets determined by elements of $\mathcal{X}_N$ itself, whereas the actualist Continuity axiom considers contour sets in $\mathcal{X}_N$ determined by *all* elements of $\mathcal{X}_\propto$.

Of course, the possibilist Continuity and Separability axioms *could* be formulated in the actualist framework and might even appear reasonably natural if we were already fully committed to a universal zero critical level. We could then formulate a version of Theorem 1 in the actualist framework, using these two axioms (together with Anonymity and Pareto), along with an axiom universal zero critical level. However, the fact that these two axioms are substantially weaker than their actualist counterparts, and yet *still* sufficient to characterize the (possibilist) rank-additive SWOs *without* the need for Neutral population growth or Trade-off Consistency, tells us that universal zero critical level is in fact a very powerful axiom—powerful enough that it is perhaps more appropriate to think of it as determining an entirely different analytical framework, rather than just a special case of the actualist framework.[21]

4 Existence independence

Rank-additive SWOs violate an axiom which Blackorby et al. (2005, §5.6) call Existence independence. This axiom says that the ethical evaluation of outcomes concerning some collection $\mathcal{K}$ of individuals (say, those currently alive on planet Earth) should not depend upon information about the lifetime utilities—or even the existence—of people outside of $\mathcal{K}$ (say, people who died long ago, who will be born in the far future, or who live on other planets). As Blackorby et al. (2005, §5.1.1) note, the ethical evaluation of presently existing people should not depend on the utility of some long-dead historical figure, such as Euclid. Likewise, suppose that a colony of humans on another planet has long ago lost all contact with Earth; Blackorby et al. (2005) argue that it would be absurd if the ethical evaluations of the colonists depended upon the utilities of the earthlings (or vice versa).[22]

The generalized utilitarian SWF in formula (2D) satisfies Existence Independence, as does any "critical-level" variant of generalized utilitarianism (with a constant critical level). But it is violated by average utilitarianism, number-dampened utilitarianism, and any other SWF where the critical level depends on the utilities of already existing people. Rank-additive SWOs violate Existence Independence in an even more

[21] Alternately, instead of imposing a universal zero critical level directly through an axiom, we could derive it as a consequence of Continuity and the conjunction of two other axioms, namely Positive expansion and Negative expansion. This shows that these two axioms are also stronger than they look.

[22] See section 4 of Thomas (2019) for further discussion of these arguments.

fundamental way: if $\mathcal{K}$ is the collection of individuals under consideration, then we don't even know how to assign *ranks* to the members of $\mathcal{K}$ until we know the lifetime utilities of all the other people not in $\mathcal{K}$. We know almost nothing about the well-being of the vast majority of people who have existed in human history (say, over the last 250,000 years). This creates problems for any SWO whose assessment of present and future social outcomes is sensitive to such historical data.[23] There are several possible responses:

(A) Interpret social outcomes in $\mathcal{X}$ as specifying only the lifetime utilities of individuals who will be *affected* by policy decisions; treat everyone else as ethically irrelevant. (In particular, ignore anyone who is already dead.)

(B) Interpret social outcomes in $\mathcal{X}$ as specifying only the lifetime utilities of individuals living in the present or the future. Ignore the past.

(C) Interpret social outcomes in $\mathcal{X}$ as specifying all individuals whose lifetime utilities are already known or can be predicted (including some people in the past). Ignore people about which nothing can be known.

(D) Interpret social outcomes in $\mathcal{X}$ as specifying only the lifetime utilities of individuals living after a fixed date (e.g. January 1, 2018). Ignore everyone before this date.

(E) Treat the utilities of unobserved individuals as a source of policy uncertainty, and deal with it the same way we deal with any other source of uncertainty: by positing a probability distribution over the unknown variables and then maximizing expected value with respect to this probability distribution.

The problem with (A) is that it is not entirely predictable who will be affected by our decisions in the future. For example, suppose the lost colony world unexpectedly re-establishes contact with Earth, after many centuries of isolation; at this moment, the rankings of everyone on the colony and on Earth would need to be recalculated, possibly leading to large changes in the evaluation of social policies. In particular, if $\mathbf{x}$ and $\mathbf{y}$ are two social outcomes which concern only the colonists and $\mathbf{x}'$ and $\mathbf{y}'$ are two social outcomes which concern only earthlings, then we may end up with a perverse situation where $\mathbf{x} \succ \mathbf{y}$ and $\mathbf{x}' \succ \mathbf{y}'$, but $\mathbf{x} \uplus \mathbf{x}' \prec \mathbf{y} \uplus \mathbf{y}'$.

Option (B) avoids this problem. But an obvious problem with both (A) and (B) is *time inconsistency*: as time passes, people move from "the future" or "the present" into "the past", and are removed from the specification of the social outcome. This changes the rankings of the remaining people, and hence, the evaluation of social outcomes. It would seem strange if social outcome $\mathbf{x}$ was deemed preferable to outcome $\mathbf{y}$ before David Bowie died, but a moment after he dies, we decide that $\mathbf{y}'$ is actually better than $\mathbf{x}'$ (where $\mathbf{x}'$ and $\mathbf{y}'$ are obtained by removing Bowie's lifetime utility from $\mathbf{x}$ and $\mathbf{y}$, respectively).

Approach (C) avoids time inconsistency. But it can still respond perversely to the arrival of new information. For example, a new and unanticipated archaeological discovery could change our estimate of the lifetime utilities of the citizens of a large

[23] This also raises the question of whether we should include proto-human species such as *Homo neanderthalensis* or *Homo heidelbergensis* in the scope of the SWO. This is a deep and fascinating philosophical problem. But by the same token, it creates even more difficulties for SWOs which violate Independence of the wretched.

ancient civilization (say, the Achaemenid Empire) and thus perturb our evaluation of social outcomes in the present day. Again, this seems absurd.

Approach (D) avoids the problems of (A), (B), and (C), but it is motivated more by pragmatism than by principle; certainly, we must give up any pretentions of moral realism if we allow our ethical evaluations to depend on an arbitrarily stipulated date on a calendar. Furthermore, (D) is still vulnerable to unknown information about the future; since we cannot really predict the lifetime utilities of far future people with any degree of precision, how are we supposed to incorporate them into the social welfare evaluation?

This leaves us with approach (E). Approach (E) does not try to exclude unknown or unknowable lifetime utilities from the specification of the social outcome by some arbitrary criterion. Instead, it "bites the bullet", acknowledging that these unknowns exist, they are ethically relevant, and they must be taken into account. To formalize approach (E), I will assume the possibilist framework of Sect. 2. (The formalization for actualist SWOs is similar and is left to the reader.)[24] Let $\mathcal{I} = \mathcal{J} \sqcup \mathcal{K}$, where $\mathcal{J}$ is an infinite set representing all potential *unobserved* individuals (living in the distant future, the forgotten past, or on faraway planets), while $\mathcal{K}$ is another infinite set representing all *observable* individuals (e.g. those presently alive on Earth). Let $\mathcal{Y} := \{\mathbf{y} \in \mathbb{R}^{\mathcal{J}}$; only finitely many coordinates of $\mathbf{y}$ are nonzero$\}$ and let $\mathcal{Z} := \{\mathbf{z} \in \mathbb{R}^{\mathcal{K}}$; only finitely many coordinates of $\mathbf{z}$ are nonzero$\}$. Then $\mathcal{X} = \mathcal{Y} \times \mathcal{Z}$. Let μ be a probability distribution over $\mathcal{Y}$, representing our beliefs about the lifetime utilities of all unobserved people. For any social outcome $\mathbf{z} \in \mathcal{Z}$ (representing the lifetime utilities of *observed* people), define

$$\widetilde{W}(\mathbf{z}) := \int_{\mathcal{Y}} W(\mathbf{y} \uplus \mathbf{z}) \ \mathrm{d}\mu[\mathbf{y}]. \tag{4A}$$

This defines a new SWF $\widetilde{W} : \mathcal{Z} \longrightarrow \mathbb{R}$, and it is *this* SWF that (E) says we should maximize. How does this work from a practical point of view? Let $\mathbf{z} := (\mathbf{z}^+, \mathbf{z}^-) \in \mathbb{R}_+^{\propto\downarrow} \times \mathbb{R}_-^{\propto\uparrow}$. For any $n \in \mathbb{N}$, we can define a probability distribution $\rho^+_{\mathbf{z},n}$ on $[n \ldots \infty)$ where $\rho^+_{\mathbf{z},n}(m)$ is the probability (according to μ) that the individual in $\mathcal{K}$ with lifetime utility z^+_n *actually* has rank m amongst all individuals with positive utility, once we take into account all the unobserved individuals in $\mathcal{J}$. Likewise, for any $n \in \mathbb{N}$, we define a probability distribution $\rho^-_{\mathbf{z},n}$ on $[n \ldots \infty)$, where $\rho^-_{\mathbf{z},n}(m)$ is the probability that the individual in $\mathcal{K}$ with lifetime utility z^-_n actually has rank m amongst all individuals with negative utilities. (Note that $\rho^{\pm}_{\mathbf{z},n}$ are only supported on $[n \ldots \infty)$, because introducing new individuals to the list can only *increase* the rank of any existing individual.) We then define the functions $\widetilde{\phi}^{\pm}_n : \mathbb{R}_{\pm} \longrightarrow \mathbb{R}_{\pm}$ by

$$\widetilde{\phi}^{\pm}_n(r) := \sum_{m=n}^{\infty} \rho^{\pm}_{\mathbf{z},n}(m)\, \phi^{\pm}_m(r), \quad \text{for all } r \in \mathbb{R}_{\pm}. \tag{4B}$$

The SWF $\widetilde{W}$ in formula (4A) is then simply the rank-additive SWF obtained by inserting $\{\widetilde{\phi}^+_n\}_{n=1}^{\infty}$ and $\{\widetilde{\phi}^-_n\}_{n=1}^{\infty}$ into formula (2B).

[24] For another rank-dependent approach to population ethics with uncertainty, see Asheim and Zuber (2016).

Of course, approach (E) faces the same question as any decision under uncertainty: How can we construct the probability distribution μ? But this question already confronts *any* social decision problem which concerns people living in the far future. One way to minimize the dependency on μ is to minimize the amount of variation between the functions $\{\phi_n^+\}_{n=1}^\infty$ and $\{\phi_n^-\}_{n=1}^\infty$—or more importantly, between their derivatives. If the derivatives $\{(\phi_n^+)'\}_{n=1}^\infty$ are all very similar to one another, then the derivatives $\{(\widetilde{\phi}_n^+)'\}_{n=1}^\infty$ of the functions defined in formula (4B) will also be very similar, independent of the precise choice of μ (and likewise for $\{(\phi_n^-)'\}_{n=1}^\infty$ and $\{(\widetilde{\phi}_n^-)'\}_{n=1}^\infty$). Proposition 2.2 tells us that the functions $\{\phi_n^+\}_{n=1}^\infty$ must rapidly decay to zero in a neighbourhood of the neutral utility 0. Hence, in this neighbourhood, we cannot expect them to be similar in this desired sense. But outside of this neighbourhood, nothing prevents us from ensuring that their derivatives are as similar as possible; see Example 2.7. If $(\phi_{100}^+)'$ and $(\phi_{10000}^+)'$ are almost the same, then it doesn't matter whether a certain individual is ranked 100th or 10,000th—the marginal social welfare contribution of her lifetime utility is almost the same in both cases, so she will treated the same in any policy decision in both cases.

This consideration suggests *avoiding* the rank-weighted utilitarian SWFs such as (2E) and (3D), where inequality aversion is obtained by systematically increasing the slopes of the functions $\{\phi_n^+\}_{n=1}^\infty$ as $n\to\infty$ (and systematically decreasing the slopes of the functions $\{\phi_n^-\}_{n=1}^\infty$ as $n\to\infty$). Instead, it suggests that we use something like the generalized utilitarian SWF in formula (2D), where the functions $\{\phi_n^+\}_{n=1}^\infty$ are all as similar as possible, and inequality aversion is obtained by making them sufficiently concave.

5 Excess (in)egalitarianism

As explained in Example 3.3, an ascending rank-weighted generalized utilitarian (ARWGU) SWO (3D) is *proper* if $\sum_{n=1}^\infty a_n < \infty$. If $\{a_n\}_{n=1}^\infty$ is decreasing, then such an SWO satisfies both Inequality aversion and No Repugnant Conclusion—an attractive combination. However, these SWOs have a problem. If there is a sufficiently large number of people with "satisfactory" lives, then a proper ARWGU SWO will prioritize the needs of a small population with slightly worse lives over the creation of an arbitrarily large population with excellent lives. To see this, note that, for any $\epsilon > 0$, there is some $N(\epsilon) \in \mathbb{N}$ such that $\sum_{k=N(\epsilon)+1}^\infty a_k < \epsilon$. Suppose for simplicity that ϕ is the identity. (The same argument works for any choice of ϕ). Consider a population $\mathbf{x}$ consisting of a large number N of people with lifetime utility 100 (representing a "satisfactory" life) and a much smaller number M of people with lifetime utility 99. For concreteness, say that $M = 50$. Let $B := \sum_{k=1}^M a_k$, let $\epsilon := B/1{,}000{,}000$, and suppose that $N > N(\epsilon)$. Now consider the following options:

- $\mathbf{y}$ consists of $N + M$ people, all having lifetime utility 100.
- $\mathbf{z} = \mathbf{x} \uplus \mathbf{u}$, where $\mathbf{u}$ is a "utopia" containing a trillion people, all having lifetime utility 1,000,000.

It is easily seen that $W(\mathbf{y}) > W(\mathbf{z})$. Formally:

$$\begin{aligned} W(\mathbf{y}) &= W(\mathbf{x}) + \sum_{n=1}^{M} a_k = W(\mathbf{x}) + B \\ &= W(\mathbf{x}) + 1{,}000{,}000\,\epsilon > W(\mathbf{x} \uplus \mathbf{u}) = W(\mathbf{z}). \end{aligned}$$

In other words, the ARWGU SWO represented by W considers it better to help 50 people slightly improve their lifetime utility from 99 to 100, rather than to create a utopia with a trillion people leading excellent lives, each with a lifetime utility of 1,000,000.

For concreteness, let's say $N = 10$ billion, and that 100 represents the lifetime utility of the average middle-class person in a Western European country in the early twenty-first century. So $\mathbf{x}$ represents a world somewhat more populous than our own, but with poverty entirely eliminated worldwide. Perhaps it is Earth two hundred years in the future. Now suppose that astronomers discover that this world faces an apocalyptic threat—say, it is about to be struck by a huge asteroid, and the resulting explosion will destroy all life on the planet. However, for a relatively small investment of resources, it would be possible to evacuate some fraction of humanity to a self-sufficient lunar colony. (This is a future where the technological problems of space travel and lunar settlement have been solved.) Let us suppose that this lunar colony will not only survive, but flourish, and give rise to a vast and long-lived interstellar civilization (represented by $\mathbf{u}$) which, over the coming millennia will be home to a trillion inhabitants who all live very long, happy, and fulfilling lives. For the sake of the thought experiment, suppose (implausibly) that this happy outcome is guaranteed in advance, and is known to the inhabitants of Earth. This is outcome $\mathbf{z}$.

Alternately, instead of saving human civilization, we could use these same resources to slightly improve the well-being of a small but unfortunate minority, who have slightly subaverage lifetime utility (i.e. 99 instead of 100). Perhaps they need minor cosmetic surgery. This is outcome $\mathbf{y}$. Most people's moral intuitions say that $\mathbf{z}$ is better than $\mathbf{y}$. But according to the ARWGU social welfare function W says $\mathbf{y}$ is better than $\mathbf{z}$.[25]

Here is another counterintuitive consequence. For any $N \in \mathbb{N}$, let $\mathbf{x}^N$ describe a world containing N million people, where the vast majority (say, 99.9999%) have excellent lives (say, a lifetime utility of 10,000) but a tiny minority (0.00001%) have lives so terrible that they are not even worth living (say, a lifetime utility of -1). Any "utopia" which one can imagine will have welfare distribution something like this: no matter how perfect the utopia, there will inevitably be some tiny fraction of people who, through simple bad luck, end up with miserable lives—perhaps they suffer from some extremely rare disease, or perhaps they are victims of some incredibly improbable but terrible accident.

[25] This is reminiscent of Fleurbaey and Tungodden's (2010) *tyranny of nonaggregation*, but it is not the same thing. Indeed, Proposition 3.4(a) showed that any ARA (including any ARWGU) satisfies Minimal Aggregation and hence avoids the tyranny of nonaggregation. But the tyranny of nonaggregation involves a fixed population, whereas the paradox presented here depends on a variable population.

One would think that such a utopia is so wonderful that we should make N as large as possible. But according to the SWO W, the larger we make N, *the worse* $\mathbf{x}^N$ *becomes*.[26] Indeed, if N is large enough, then total social welfare is *negative*, meaning that a vast galactic utopia with the above statistical welfare distribution of well-being is ethically *worse* than a totally lifeless galaxy. For a less stark comparison, let $\mathbf{y}$ be a "small, safe, but boring" world, containing only one million people, all of whom have lives which are wretched, but technically worth living (say, a lifetime utility of 1). It is easily verified that, if N is large enough, then $W(\mathbf{x}^N) < W(\mathbf{y})$. Suppose humanity had to choose between two futures: one leading to a galactic utopia ($\mathbf{x}^N$, for large N) and the other leading to a wretched but anodyne future ($\mathbf{y}$). The SWO W says that humanity should choose $\mathbf{y}$.

Excess egalitarianism in particular affects the *rank-discounted utilitarian* SWO (3E) characterized by Asheim and Zuber (2014). However, if proper ARWGU SWOs suffer from excess egalitarianism, then possibilist RA SWOs can suffer from an even worse problem: excess *in*egalitarianism. To see this, recall from Proposition 2.2(a) that a possibilist RA SWO with SWF W as in (2B) satisfies No Repugnant Conclusion if and only if there exists $r_0 > 0$ such that $\sum_{n=1}^{\infty} \phi_n^+(r_0) < \infty$. Thus, for any $\epsilon > 0$, there is some $N(\epsilon)$ such that $\sum_{n=N+1}^{\infty} \phi_n^+(r_0) < \epsilon$.

For concreteness, suppose that $r_0 = 1$, while a lifetime utility level 100 represents, say, a middle-class life in a Western European country. Let $\epsilon := 0.0001 \times (\phi_1^+)'(100)$; this is roughly the increase in total value that would be obtained if the best-off person in society increased her lifetime utility from 100 to 100.0001. Let $N := N(\epsilon)$, and let $M := 1{,}000{,}000{,}000\,N$. Let $\mathbf{x}$ be a social outcome containing N "well-off" people with lifetime utility 100, and M "miserable" people with a lifetime utility of 0.1 (that is, lives of abject misery, barely worth living). Consider the following possible improvements:

- In $\mathbf{y}$, the best-off person's lifetime utility is increased from 100 to 100.0002, while the lifetime utility of everyone else stays exactly the same as in $\mathbf{x}$.
- In $\mathbf{z}$, the N well-off people remain the same, while the lifetime utilities of the M miserable people are increased from 0.1 to 1.

It is easily verified that $W(\mathbf{y}) > W(\mathbf{z})$; in other words, the SWO considers it better to increase the utility of the most fortunate individual by a minuscule amount, rather than significantly boost the utilities of an astronomically vast population of miserable people.

6 Conclusion

Excess egalitarianism and excess inegalitarianism are very unappealing problems, which plague any rank-additive SWO (either actualist or possibilist) that avoids the Repugnant Conclusion via Propositions 2.2(a) and 3.2(a). In the light of this, Theorems 1 and 2 might not seem like positive results, but rather impossibility theorems. Rank-additive SWOs also have other shortcomings: actualist SWOs violate Positive

[26] This illustrates that ARAs are not replication-invariant, as already noted in Sect. 3.3.

and Negative expansion, while possibilist SWOs violate Inequality aversion. As always in population ethics, there are trade-offs to be made. What is the best way to make them? This is an interesting problem for future research.

A Appendix: Proofs from Section 2

Proof of Theorem 1 The proof of "$\Longleftarrow$" is straightforward, so I will focus on the proof of "$\Longrightarrow$". First I will show that each of the orders $\succeq_N$ admits an additive representation on $\mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow}$. Then I will combine all these representations together to obtain a rank-additive SWF on $\mathbb{R}_+^{\propto\downarrow} \times \mathbb{R}_-^{\propto\uparrow}$. To achieve the first of these steps, I will combine the classic representation theorem of Debreu (1960) with a well-known result of Chateauneuf and Wakker (1993) (see Claim 6). But the deployment of this result requires some technical preliminaries; this is the role of Claims 1–5.

For any $N \in \mathbb{N}$, let $\mathbb{R}_{++}^{N\downarrow\downarrow} := \{\mathbf{x} \in \mathbb{R}^N;\ x_1 > x_2 > \cdots > x_N > 0\}$ and $\mathbb{R}_{--}^{N\uparrow\uparrow} := \{\mathbf{x} \in \mathbb{R}^N;\ x_1 < x_2 < \cdots < x_N < 0\}$. Clearly, $\mathbb{R}_{++}^{N\downarrow\downarrow}$ is the topological interior of $\mathbb{R}_+^{N\downarrow}$ as a subset of $\mathbb{R}^N$, while $\mathbb{R}_{--}^{N\uparrow\uparrow}$ is the topological interior of $\mathbb{R}_-^{N\uparrow}$. Thus, $\mathbb{R}_{++}^{N\downarrow\downarrow} \times \mathbb{R}_{--}^{N\uparrow\uparrow}$ is the interior of $\mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow}$ in $\mathbb{R}^{2N}$.

Claim 1 *Let $N \in \mathbb{N}$. Every indifference set of $\succeq_N$ in $\mathbb{R}_{++}^{N\downarrow\downarrow} \times \mathbb{R}_{--}^{N\uparrow\uparrow}$ is connected.*

Before proving Claim 1, we must develop some machinery. For any $N \in \mathbb{N}$ and any $\mathbf{r} = (r_1, \ldots, r_N) \in \mathbb{R}^N$, let $\|\mathbf{r}\| := \sqrt{r_1^2 + \cdots + r_N^2}$ be its Euclidean norm. For any $\mathbf{x} = (\mathbf{x}^+, \mathbf{x}^-)$ in $\mathbb{R}_{++}^{N\downarrow\downarrow} \times \mathbb{R}_{--}^{N\uparrow\uparrow}$, define

$$\langle\!|\mathbf{x}|\!\rangle := \sqrt{\|\mathbf{x}^+\|^2 + \frac{1}{\|\mathbf{x}^-\|^2}}.$$

(This is always well defined because $\|\mathbf{x}^-\| \neq 0$ for all $\mathbf{x} \in \mathbb{R}_{--}^{N\uparrow\uparrow}$). As the notation suggests, this will be like a sort of "pseudo-norm" on $\mathbb{R}_{++}^{N\downarrow\downarrow} \times \mathbb{R}_{--}^{N\uparrow\uparrow}$ (even though it is not a norm). For any $r \in \mathbb{R}_{++}$ and $\mathbf{x} \in \mathbb{R}_{++}^{N\downarrow\downarrow} \times \mathbb{R}_{--}^{N\uparrow\uparrow}$, we define $r \star \mathbf{x} := (r\,\mathbf{x}^+, \frac{1}{r}\mathbf{x}^-)$. It is easily verified that $\langle\!|r \star \mathbf{x}|\!\rangle = r\,\langle\!|\mathbf{x}|\!\rangle$. Let $\mathcal{S}^N := \{\mathbf{s} \in \mathbb{R}_{++}^{N\downarrow\downarrow} \times \mathbb{R}_{--}^{N\uparrow\uparrow};\ \langle\!|\mathbf{s}|\!\rangle = 1\}$; this plays the role of the "unit sphere" for this "norm". For any $\mathbf{x} \in \mathbb{R}_{++}^{N\downarrow\downarrow} \times \mathbb{R}_{--}^{N\uparrow\uparrow}$, if $r := \langle\!|\mathbf{x}|\!\rangle$, then $\frac{1}{r} \star \mathbf{x} \in \mathcal{S}^N$.

Claim 2 *Let $N \in \mathbb{N}$, and let $\mathbf{x} \in \mathbb{R}_{++}^{N\downarrow\downarrow} \times \mathbb{R}_{--}^{N\uparrow\uparrow}$. Let $\mathcal{Z} := \{\mathbf{z} \in \mathbb{R}_{++}^{N\downarrow\downarrow} \times \mathbb{R}_{--}^{N\uparrow\uparrow};\ \mathbf{z} \approx_N \mathbf{x}\}$ be the indifference set of $\mathbf{x}$. For any $\mathbf{s} \in \mathcal{S}^N$, there is a unique $r \in \mathbb{R}_{++}$ with $r \star \mathbf{s} \in \mathcal{Z}$. Let $\phi(\mathbf{s}) := r \star \mathbf{s}$; this defines a continuous surjection $\phi : \mathcal{S}^N \longrightarrow \mathcal{Z}$.*

Proof *Existence and uniqueness* Since $\mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow}$ is a connected, separable topological space, and $\succeq_N$ satisfies Continuity, the theorem of Debreu (1954) yields a continuous function $w : \mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow} \longrightarrow \mathbb{R}$ that represents $\succeq_N$—i.e. for all

$(\mathbf{a}^+, \mathbf{a}^-), (\mathbf{b}^+, \mathbf{b}^-) \in \mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow}$, we have $(\mathbf{a}^+, \mathbf{a}^-) \succeq (\mathbf{b}^+, \mathbf{b}^-)$ if and only if $w(\mathbf{a}^+, \mathbf{a}^-) \geqslant w(\mathbf{b}^+, \mathbf{b}^-)$. Furthermore, w is increasing in every coordinate, because $\succeq_N$ satisfies Pareto.

Fix $\mathbf{s} \in \mathcal{S}$. For any $r \in \mathbb{R}_{++}$, let $v(r) := w(r \star \mathbf{s})$. Then $v : \mathbb{R}_{++} \longrightarrow \mathbb{R}$ is clearly a continuous function. Suppose r is large enough that every coordinate of $r\mathbf{s}^+$ is larger than the corresponding coordinate of $\mathbf{x}^+$, while every coordinate of $\frac{1}{r}\mathbf{s}^-$ is smaller in magnitude than the corresponding coordinate of $\mathbf{x}^-$. Then $r \star \mathbf{s} \succ \mathbf{x}$ by Pareto, and thus, $v(r) = w(r \star \mathbf{s}) > w(\mathbf{x})$.

On the other hand, suppose r is small enough that every coordinate of $r\mathbf{s}^+$ is less than the corresponding coordinate of $\mathbf{x}^+$, while every coordinate of $\frac{1}{r}\mathbf{s}^-$ is larger in magnitude than the corresponding coordinate of $\mathbf{x}^-$. Then $r \star \mathbf{s} \prec \mathbf{x}$ by Pareto, and thus, $v(r) = w(r \star \mathbf{s}) < w(\mathbf{x})$.

Since w is continuous, the intermediate value theorem yields some $r \in \mathbb{R}_{++}$ such that $v(r) = w(\mathbf{x})$—in other words, $w(r \star \mathbf{s}) = w(\mathbf{x})$, and hence $r \star \mathbf{s} \approx_N \mathbf{x}$. Thus, $r \star \mathbf{s} \in \mathcal{Z}$, as desired. This proves that such a r exists. The fact that it is unique follows from the Pareto axiom. This argument works for all $\mathbf{s} \in \mathcal{S}$.

Surjective Given $\mathbf{z} \in \mathcal{Z}$, let $r := \langle\!|\mathbf{z}|\!\rangle$ and let $\mathbf{s} := \frac{1}{r} \star \mathbf{z}$; then, $\mathbf{s} \in \mathcal{S}^N$. But $r \star \mathbf{s} = \mathbf{z}$. Thus, $r \star \mathbf{s} \in \mathcal{Z}$, so $\phi(\mathbf{s}) = r \star \mathbf{s} = \mathbf{z}$.
Continuity For any $\mathbf{s} \in \mathcal{S}$ and $\delta > 0$, let $\mathcal{B}(\mathbf{s}, \delta) := \{\mathbf{b} \in \mathcal{S}^N;\ \|\mathbf{b} - \mathbf{s}\| < \delta\}$. For any $\epsilon > 0$, we will find a $\delta > 0$ such that $\|\phi(\mathbf{b}) - \phi(\mathbf{s})\| < \epsilon$ for all $\mathbf{b} \in \mathcal{B}(\mathbf{s}, \delta)$.

Suppose that $\phi(\mathbf{s}) = r_0 \star \mathbf{s}$ for some $r_0 \in \mathbb{R}_{++}$. For any $\mathbf{b} = (\mathbf{b}^+, \mathbf{b}^-) \in \mathcal{S}^N$, define

$$\overline{R}(\mathbf{b}) := r_0 \cdot \max\left\{\frac{s_1^+}{b_1^+}, \ldots, \frac{s_N^+}{b_N^+}, \frac{b_1^-}{s_1^-}, \ldots, \frac{b_N^-}{s_N^-}\right\}.$$

If $r > \overline{R}(\mathbf{b})$, then $r\, b_n^+ > r_0\, s_n^+$ and $\frac{1}{r}\, b_n^- > \frac{1}{r_0}\, s_n^-$ for all $n \in [1 \ldots N]$; thus, $r \star \mathbf{b} = (r\, \mathbf{b}^+, \frac{1}{r}\mathbf{b}^-) \succ (r_0\, \mathbf{s}^+, \frac{1}{r_0}\mathbf{s}^-) = \phi(\mathbf{s}) \approx \mathbf{z}$, so that $r \star \mathbf{b} \notin \mathcal{Z}$. (Here, the "$\succ$" is by Pareto.) Likewise, define

$$\underline{R}(\mathbf{b}) := r_0 \cdot \min\left\{\frac{s_1^+}{b_1^+}, \ldots, \frac{s_N^+}{b_N^+}, \frac{b_1^-}{s_1^-}, \ldots, \frac{b_N^-}{s_N^-}\right\}.$$

If $r < \underline{R}(\mathbf{b})$, then $r\, b_n^+ < r_0\, s_n^+$ and $\frac{1}{r}\, b_n^- < \frac{1}{r_0}\, s_n^-$ for all $n \in [1 \ldots N]$; thus, $r \star \mathbf{b} = (r\, \mathbf{b}^+, \frac{1}{r}\mathbf{b}^-) \prec (r_0\, \mathbf{s}^+, \frac{1}{r_0}\mathbf{s}^-) = \phi(\mathbf{s}) \approx \mathbf{z}$, so that $r \star \mathbf{b} \notin \mathcal{Z}$. (Again, the "$\prec$" is by Pareto.) Thus,

$$\phi(\mathbf{b}) \quad = \quad r \star \mathbf{b} \ \text{ for some } r \in \mathbb{R}_{++} \text{ with } \underline{R}(\mathbf{b}) < r < \overline{R}(\mathbf{b}). \tag{A1}$$

Let $\overline{\delta} := \min\{|s_n^\pm|\}_{n=1}^N$. For $\delta \in (0, \overline{\delta})$, define

$$\overline{R}(\delta) := r_0 \cdot \max\left\{\frac{s_1^+}{s_1^+ - \delta}, \ldots, \frac{s_N^+}{s_N^+ - \delta}, \frac{s_1^- - \delta}{s_1^-}, \ldots, \frac{s_N^- - \delta}{s_N^-}\right\}$$

$$\text{and} \quad \underline{R}(\delta) := r_0 \cdot \min\left\{ \frac{s_1^+}{s_1^+ + \delta}, \ldots, \frac{s_N^+}{s_N^+ + \delta}, \frac{s_1^- + \delta}{s_1^-}, \ldots, \frac{s_N^- + \delta}{s_N^-} \right\}.$$

(Note: $\delta < \overline{\delta}$, so $s_n^+ - \delta > 0$ and $s_n^- + \delta < 0$ for all $n \in [1 \ldots N]$.) Then

$$\underline{R}(\delta) \leqslant \underline{R}(\mathbf{b}) \leqslant \overline{R}(\mathbf{b}) \leqslant \overline{R}(\delta), \quad \text{for all } \mathbf{b} \in \mathcal{B}(\mathbf{s}, \delta). \tag{A2}$$

Furthermore, note that

$$\lim_{\delta \to 0} \overline{R}(\delta) = \lim_{\delta \to 0} \underline{R}(\delta) = r_0. \tag{A3}$$

Let $M := \|\mathbf{s}\| + 1$. Then

$$\|\mathbf{b}^{\pm}\| < \|\mathbf{b}\| \leqslant \|\mathbf{s}\| + 1 = M, \qquad \text{for all } \mathbf{b} = (\mathbf{b}^+, \mathbf{b}^-) \text{ in } \mathcal{B}(\mathbf{s}, 1). \tag{A4}$$

Given any $\epsilon > 0$, let $\eta > 0$ be small enough that

$$\sqrt{\eta^2 + \left(\frac{\eta}{r_0\,(r_0 - \eta)} \right)^2} < \frac{\epsilon}{2M}. \tag{A5}$$

By statement (A3), there exists some $\delta_1 \in (0, \overline{\delta})$ such that

$$|\overline{R}(\delta) - r_0| < \eta \ \text{ and } \ |\underline{R}(\delta) - r_0| < \eta, \quad \text{for all } \delta < \delta_1. \tag{A6}$$

Meanwhile, let

$$\delta_2 := \frac{\epsilon}{2\sqrt{r_0^2 + \frac{1}{r_0^2}}}. \tag{A7}$$

Finally, define $\delta := \min\{1, \delta_1, \delta_2\}$. Now, let $\mathbf{b} \in \mathcal{B}(\mathbf{s}, \delta)$, and suppose $\phi(\mathbf{b}) = r \star \mathbf{b}$ for some $r \in \mathbb{R}_{++}$. Then

$$\begin{aligned}
\|\phi(\mathbf{b}) - \phi(\mathbf{s})\| &= \|r \star \mathbf{b} - r_0 \star \mathbf{s}\| \leqslant \|r \star \mathbf{b} - r_0 \star \mathbf{b}\| + \|r_0 \star \mathbf{b} - r_0 \star \mathbf{s}\| \\
&= \sqrt{|r - r_0|^2 \left\|\mathbf{b}^+\right\|^2 + \left|\frac{1}{r} - \frac{1}{r_0}\right|^2 \left\|\mathbf{b}^-\right\|^2} + \sqrt{r_0^2 \left\|\mathbf{b}^+ - \mathbf{s}^+\right\|^2 + \frac{1}{r_0^2}\left\|\mathbf{b}^- - \mathbf{s}^-\right\|^2} \\
&\underset{(a)}{\leqslant} M \sqrt{|r - r_0|^2 + \left|\frac{1}{r} - \frac{1}{r_0}\right|^2} + \sqrt{r_0^2\,\delta^2 + \frac{1}{r_0^2}\,\delta^2} \\
&\underset{(b)}{\leqslant} M \sqrt{\eta^2 + \left(\frac{\eta}{r_0\,(r_0 - \eta)}\right)^2} + \delta \sqrt{r_0^2 + \frac{1}{r_0^2}} \\
&\underset{(c)}{\leqslant} M \sqrt{\eta^2 + \left(\frac{\eta}{r_0\,(r_0 - \eta)}\right)^2} + \delta_2 \sqrt{r_0^2 + \frac{1}{r_0^2}} \underset{(d)}{\leqslant} \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,
\end{aligned}$$

as desired. Here, (a) is because $\|\mathbf{b}^+ - \mathbf{s}^+\| < \delta$ and $\|\mathbf{b}^- - \mathbf{s}^-\| < \delta$ because $\mathbf{b} \in \mathcal{B}(\mathbf{s}, \delta)$, while $\|\mathbf{b}^\pm\| \leqslant M$, by inequality (A4), because $\delta \leqslant 1$. Next, (b) is because $r \in (r_0 - \eta, r_0 + \eta)$ by statements (A1), (A2), and (A6), because $\mathbf{b} \in \mathcal{B}(\mathbf{s}, \delta)$ and $\delta \leqslant \delta_1$. Meanwhile, (c) is because $\delta \leqslant \delta_2$. Finally, (d) is by definitions (A5) and (A7). ◇ Claim 2

Claim 3 *For any $N \in \mathbb{N}$, $\mathcal{S}^N$ is path-connected.*

Proof For any $r \in (0, 1)$, let

$$\mathcal{S}^N_+(r) \;:=\; \left\{\mathbf{x} \in \mathbb{R}^{N\downarrow\downarrow}_{++} \;;\; \|\mathbf{x}\| = r\right\} \;\text{ and }\; \mathcal{S}^N_-(r) \;:=\; \left\{\mathbf{x} \in \mathbb{R}^{N\uparrow\uparrow}_{--} \;;\; \|\mathbf{x}\| = r\right\}.$$

Then it is easily verified that

$$\mathcal{S}^N := \bigsqcup_{r\in(0,1)} \left(\mathcal{S}^N_+(r) \times \mathcal{S}^N_-\left(\frac{1}{\sqrt{1-r^2}}\right)\right). \tag{A8}$$

Now let $\mathbf{p} = (\mathbf{p}^+, \mathbf{p}^-)$ and $\mathbf{r} = (\mathbf{r}^+, \mathbf{r}^-)$ be two elements of $\mathcal{S}^N$. Let $p_+ := \|\mathbf{p}^+\|$ and $r_+ := \|\mathbf{r}^+\|$, and let $p_- := 1/\sqrt{1-p_+^2}$ and $r_- := 1/\sqrt{1-r_+^2}$. Then equation (A8) implies that $\mathbf{p} \in \mathcal{S}^N_+(p_+) \times \mathcal{S}^N_-(p_-)$ and $\mathbf{r} \in \mathcal{S}^N_+(r_+) \times \mathcal{S}^N_-(r_-)$. Now, define

$$\mathbf{q}^+ := \frac{p_+}{r_+}\mathbf{r}^+ \quad\text{and}\quad \mathbf{q}^- := \frac{p_-}{r_-}\mathbf{r}^-.$$

Then $\mathbf{q}^+ \in \mathbb{R}^{\propto\downarrow\downarrow}_{++}$ and $\mathbf{q}^- \in \mathbb{R}^{\propto\uparrow\uparrow}_{--}$ (because $\mathbf{r}^+ \in \mathbb{R}^{\propto\downarrow\downarrow}_{++}$ and $\mathbf{r}^- \in \mathbb{R}^{\propto\uparrow\uparrow}_{--}$) and $\|\mathbf{q}^+\| = p_+$ and $\|\mathbf{q}^-\| = p_-$. Thus, if $\mathbf{q} := (\mathbf{q}^+, \mathbf{q}^-)$ then $\mathbf{q} \in \mathcal{S}^N_+(p_+) \times \mathcal{S}^N_-(p_-)$; hence, $\mathbf{q} \in \mathcal{S}^N$.

Now $\mathcal{S}^N_+(p_+)$ is path-connected, since it is the intersection of the convex cone $\mathbb{R}^{\propto\downarrow\downarrow}_{++}$ with the radius-p_+ sphere around 0 in $\mathbb{R}^N$. Likewise, $\mathcal{S}^N_-(p_-)$ is path-connected. Thus, the Cartesian product $\mathcal{S}^N_+(p_+) \times \mathcal{S}^N_-(p_-)$ is also path-connected. Thus, there is a continuous function $\gamma : [-1, 0] \longrightarrow \mathcal{S}^N_+(p_+) \times \mathcal{S}^N_-(p_-)$ such that $\gamma(-1) = \mathbf{p}$ and $\gamma(0) = \mathbf{q}$. Next, for all $t \in [0, 1]$, let $\rho_+(t) := t\, r_+ + (1-t)\, p_+$, and define $\rho_-(t) := 1/\sqrt{1-\rho_+(t)^2}$. Then $\rho^\pm : [0, 1] \longrightarrow (0, 1)$ are continuous functions, with $\rho_+(0) = p_+$ and $\rho_-(0) = p_-$, while $\rho_+(1) = r_+$ and $\rho_-(1) = r_-$. Define $\gamma : [0, 1] \longrightarrow \mathcal{S}^N$ by

$$\gamma(t) := \left(\frac{\rho_+(t)}{r_+}\mathbf{r}^+, \; \frac{\rho_-(t)}{r_-}\mathbf{r}^-\right), \quad \text{for all } t \in [0, 1].$$

Then γ is a continuous function, with $\gamma(0) = \mathbf{q}$ and $\gamma(1) = \mathbf{r}$. Furthermore, $\gamma(t) \in \mathcal{S}^N$ for all $t \in [0, 1]$ by equation (A8).

At this point, we have constructed a continuous function $\gamma : [-1, 1] \longrightarrow \mathcal{S}^N$ such that $\gamma(-1) = \mathbf{p}$ and $\gamma(1) = \mathbf{r}$. This works for any $\mathbf{p}, \mathbf{r} \in \mathcal{S}^N$. Thus, $\mathcal{S}^N$ is connected. ◇ Claim 3

Proof of Claim 1 Let $\mathcal{Z}$ be an indifference set of $\succeq_N$ in $\mathbb{R}^{N\downarrow\downarrow}_{++} \times \mathbb{R}^{N\uparrow\uparrow}_{--}$. Claim 2 says that $\mathcal{Z}$ is the image of $\mathcal{S}^N$ under a continuous function. Claims 3 says $\mathcal{S}^N$ is path-connected. The continuous image of a path-connected set is also connected. Thus, $\mathcal{Z}$ is path-connected. $\diamond$ Claim 1

Claim 4 *For every* $\mathbf{x} \in \mathbb{R}^{N\downarrow}_{+} \times \mathbb{R}^{N\uparrow}_{-}$, *there is some* $\mathbf{y} \in \mathbb{R}^{N\downarrow\downarrow}_{++} \times \mathbb{R}^{N\uparrow\uparrow}_{--}$ *such that* $\mathbf{x} \approx_N \mathbf{y}$.

Proof As explained at the start of the proof of Claim 2, there is a continuous function $w : \mathbb{R}^{N\downarrow}_{+} \times \mathbb{R}^{N\uparrow}_{-} \longrightarrow \mathbb{R}$ that is increasing in every coordinate and that represents $\succeq_N$. Suppose $\mathbf{x} = (\mathbf{x}^+, \mathbf{x}^-)$. Let $\mathbf{z}^+ \in \mathbb{R}^{N\downarrow\downarrow}_{++}$ be obtained by increasing all coordinates of $\mathbf{x}^+$ slightly, so that $z_1^+ > z_2^+ > \cdots > z_N^+ > 0$. Thus, $w(\mathbf{z}^+, \mathbf{x}^-) > w(\mathbf{x})$, by Pareto. Let $\mathbf{z}^- \in \mathbb{R}^{N\uparrow}_{-}$ be obtained by decreasing all coordinates of $\mathbf{x}^-$ slightly, so that $z_1^- < z_2^- < \cdots < z_N^- < 0$. Thus, $w(\mathbf{x}^+, \mathbf{z}^-) < w(\mathbf{x})$, by Pareto. Now, for all $r \in [0, 1]$, let $\mathbf{y}^+(r) := r\,\mathbf{z}^+ + (1-r)\,\mathbf{x}^+$ and let $\mathbf{y}^-(r) := r\,\mathbf{x}^- + (1-r)\,\mathbf{z}^-$, and let $\mathbf{y}(r) := (\mathbf{y}^+(r), \mathbf{y}^-(r))$. Thus, $\mathbf{y}(0) = (\mathbf{x}^+, \mathbf{z}^-)$ and $\mathbf{y}(1) = (\mathbf{z}^+, \mathbf{x}^-)$. It is easily verified that $\mathbf{y}^+(r) \in \mathbb{R}^{N\downarrow\downarrow}_{++}$ for all $r \in (0, 1]$, and $\mathbf{y}^-(r) \in \mathbb{R}^{N\uparrow\uparrow}_{--}$ for all $r \in [0, 1)$; thus, $\mathbf{y}(r) \in \mathbb{R}^{N\downarrow\downarrow}_{++} \times \mathbb{R}^{N\uparrow\uparrow}_{--}$ for all $r \in (0, 1)$. Now, $w[\mathbf{y}(0)] = w(\mathbf{x}^+, \mathbf{z}^-) < w(\mathbf{x}) < w(\mathbf{z}^+, \mathbf{x}^-) = w[\mathbf{y}(1)]$, and the function $r \mapsto w[\mathbf{y}(r)]$ is clearly continuous. Thus, the intermediate value theorem yields some $r \in (0, 1)$ such that $w[\mathbf{y}(r)] = w(\mathbf{x})$. In other words $\mathbf{y}(r) \approx_N \mathbf{x}$. Now set $\mathbf{y} := \mathbf{y}(r)$ to prove the claim. $\diamond$ Claim 4

Let $\mathbf{x} = (\mathbf{x}^+, \mathbf{x}^-) \in \mathbb{R}^{N\downarrow}_{+} \times \mathbb{R}^{N\uparrow}_{-}$. For all $n \in [1 \ldots N]$, say that the coordinate x_n^+ is *interior* if there is some $\mathbf{y} \in \mathbb{R}^{N\downarrow\downarrow}_{++} \times \mathbb{R}^{N\uparrow\uparrow}_{--}$ such that $x_n^+ = y_n^+$. (Recall that $\mathbb{R}^{N\downarrow\downarrow}_{++} \times \mathbb{R}^{N\uparrow\uparrow}_{--}$ is the interior of $\mathbb{R}^{N\downarrow}_{+} \times \mathbb{R}^{N\uparrow}_{-}$ in $\mathbb{R}^{2N}$.) We likewise define the *interior* property for the coordinates $x_1^-, \ldots x_N^-$. In the terminology of Chateauneuf and Wakker (1993), $\mathbf{x}$ is *interior-matched* if $\mathbf{x} \approx_N \mathbf{y}$ for some $\mathbf{y} \in \mathbb{R}^{N\downarrow\downarrow}_{++} \times \mathbb{R}^{N\uparrow\uparrow}_{--}$ and at most one of the coordinates $x_1^+, \ldots x_N^+, x_1^-, \ldots x_N^-$ is *not* interior.[27] (Observe that the first half of this condition is automatically satisfied, by Claim 4.) Next, $\mathbf{x}$ is *second-order interior-matched* if $\mathbf{x} \approx_N \mathbf{y}$ for some interior or interior-matched $\mathbf{y} \in \mathbb{R}^{N\downarrow}_{+} \times \mathbb{R}^{N\uparrow}_{-}$, and at most one of the coordinates $x_1^+, \ldots x_N^+, x_1^-, \ldots x_N^-$ does not occur in an interior or interior-matched element. Likewise, $\mathbf{x}$ is *third-order interior-matched* if $\mathbf{x} \approx_N \mathbf{y}$ for some interior, interior-matched, or second-order interior-matched $\mathbf{y} \in \mathbb{R}^{N\downarrow}_{+} \times \mathbb{R}^{N\uparrow}_{-}$, and at most one of the coordinates $x_1^+, \ldots x_N^+, x_1^-, \ldots x_N^-$ does not occur in an interior, interior-matched element, or second-order interior-matched element. We likewise define *nth-order interior-matched* for all $n \in [1 \ldots N+1]$. Finally, $\mathbf{x}$ is *matched* if it is interior or is nth-order interior-matched for some $n \in [1 \ldots N+1]$.

Claim 5 *Every element of* $\mathbb{R}^{N\downarrow}_{+} \times \mathbb{R}^{N\uparrow}_{-}$ *is matched.*

Proof $\mathbf{x} = (\mathbf{x}^+, \mathbf{x}^-) \in \mathbb{R}^{N\downarrow}_{+} \times \mathbb{R}^{N\uparrow}_{-}$. Claim 4 guarantees that $\mathbf{x} \approx_N \mathbf{y}$ for some $\mathbf{y} \in \mathbb{R}^{N\downarrow\downarrow}_{++} \times \mathbb{R}^{N\uparrow\uparrow}_{--}$. It remains to check the matching condition on the coordinates.

[27] Actually our definition is slightly stronger than that of Chateauneuf and Wakker (1993). But it is sufficient for our purposes.

For all $n \in [1 \ldots N]$, it is easily verified that x_n^+ is interior if and only if $x_n^+ > 0$. Likewise, x_n^- is interior if and only if $x_n^- < 0$. Thus, $\mathbf{x}$ is interior-matched if and only if at most one of the coordinates $x_1^+, \ldots x_N^+, x_1^-, \ldots x_N^-$ is zero. It is easily seen that this occurs if and only if $x_{N-1}^+ > 0$ and $x_{N-1}^- < 0$, and at least one of x_N^+ and x_N^- is nonzero.

Now suppose that both $x_N^+ = 0$ and $x_N^- = 0$. Then each of these two coordinates can be matched to an interior-matched point (by the previous paragraph). Thus, in this case, $\mathbf{x}$ is second-order interior-matched if and only if all the coordinates $x_1^\pm, \ldots, x_{N-2}^\pm$ are nonzero, and at least one of the coordinates x_{N-1}^+ and x_{N-1}^- is nonzero.

If both $x_{N-1}^+ = 0$ and $x_{N-1}^- = 0$ (and hence, $x_N^+ = 0$ and $x_N^- = 0$), then each of the two coordinates x_{N-1}^+ and x_{N-1}^- can individually be matched to some second-order interior-matched point (by the previous paragraph), while each of the two coordinates x_N^+ and x_N^- can individually be matched to some interior-matched point. Thus, in this case, $\mathbf{x}$ is third-order interior-matched if and only if all the coordinates $x_1^\pm, \ldots, x_{N-3}^\pm$ are nonzero, and at least one of the coordinates x_{N-2}^+ and x_{N-2}^- is nonzero.

Proceeding inductively, we see that, for all $n \in [1 \ldots N]$, $\mathbf{x}$ is nth-order interior-matched if and only if all the coordinates $x_1^\pm, \ldots, x_{N-n}^\pm$ are nonzero, and at most one of the coordinates x_{N-n+1}^+ and x_{N-n+1}^- is zero. In particular, $\mathbf{x}$ is Nth-order interior-matched if and only if at least one of x_1^+ and x_1^- is nonzero—in other words, as long as $\mathbf{x}$ itself is not the zero vector. Thus, the zero vector itself is $(N+1)$th-order interior-matched. Hence, *every* element of $\mathbb{R}_+^{N_\downarrow} \times \mathbb{R}_-^{N_\uparrow}$ is either interior or nth-order interior-matched for some $n \in [1 \ldots N+1]$ and thus matched. ◇ Claim 5

Claim 6 *For all $N \in \mathbb{N}$ with $N \geqslant 3$, there exists a unique system of continuous, increasing functions $\psi_1^+, \ldots, \psi_N^+ : \mathbb{R}_+ \longrightarrow \mathbb{R}$ and $\psi_1^-, \ldots, \psi_N^- : \mathbb{R}_- \longrightarrow \mathbb{R}$ with $\psi_1^+(1) = 1$ and with $\psi_n^\pm(0) = 0$ for all $n \in [1 \ldots N]$, such that, for any $\mathbf{x} = (\mathbf{x}^+, \mathbf{x}^-)$ and $\mathbf{y} = (\mathbf{y}^+, \mathbf{y}^-)$ in $\mathbb{R}_+^{N_\downarrow} \times \mathbb{R}_-^{N_\uparrow}$, we have*

$$\left(\mathbf{x} \succeq_N \mathbf{y}\right) \iff \left(\sum_{n=1}^{N} \psi_n^+(x_n^+) + \sum_{n=1}^{N} \psi_n^-(x_n^-) \geqslant \sum_{n=1}^{N} \psi_n^+(y_n^+) + \sum_{n=1}^{N} \psi_n^-(y_n^-)\right). \tag{A9}$$

Proof An *open box* in $\mathbb{R}^{2N}$ is an open set of the form $(a_1, z_1) \times (a_2, z_2) \times \cdots \times (a_{2N}, z_{2N}) \subset \mathbb{R}^{2N}$, for some $a_1 < z_1, a_2 < z_2, \ldots, a_{2N} < z_{2N}$. Let $\mathcal{B} \subset \mathbb{R}_{++}^{N_{\downarrow\downarrow}} \times \mathbb{R}_{--}^{N_{\uparrow\uparrow}}$ be an open box, and let $\succeq_{\mathcal{B}}$ be the restriction of $\succeq_N$ to an ordering on $\mathcal{B}$. In the terminology of Debreu (1960), $\succeq_{\mathcal{B}}$ is continuous, separable, and increasing in every coordinate, by the axioms Continuity, Separability, and Pareto, respectively. Thus, Theorem 3 of Debreu (1960) says that $\succeq_{\mathcal{B}}$ admits an *additive representation*—that is, there are continuous, increasing functions $\psi_n^{\mathcal{B}} : (a_n, z_n) \longrightarrow \mathbb{R}$ for all $n \in [1 \ldots 2N]$ such that, for any $\mathbf{b}, \mathbf{c} \in \mathcal{B}$, we have

$$\left(\mathbf{b} \succeq_{\mathcal{B}} \mathbf{c}\right) \iff \left(\sum_{n=1}^{2N} \psi_n^{\mathcal{B}}(b_n) \geqslant \sum_{n=1}^{2N} \psi_n^{\mathcal{B}}(c_n)\right). \tag{A10}$$

$\mathbb{R}_{++}^{N\downarrow\downarrow} \times \mathbb{R}_{--}^{N\uparrow\uparrow}$ is open, so it can be covered by such open boxes. Thus, in the terminology of Chateauneuf and Wakker (1993), the ordering $\succeq_N$ admits "local" additive representations everywhere on $\mathbb{R}_{++}^{N\downarrow\downarrow} \times \mathbb{R}_{--}^{N\uparrow\uparrow}$. Since $\mathbb{R}_{++}^{N\downarrow\downarrow} \times \mathbb{R}_{--}^{N\uparrow\uparrow}$ is a convex set, it clearly satisfies conditions (1) and (2) in Structural Assumption 2.1 of Chateauneuf and Wakker (1993). Meanwhile, condition (3) of Chateauneuf and Wakker (1993) is true by Claim 1. Finally, Claim 5 says that every element of $\mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow}$ is "matched". Thus, by Theorem 3.3(a) of Chateauneuf and Wakker (1993), the local additive representations (A10) can be combined together to yield a single *global* additive representation of $\succeq_N$ on all of $\mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow}$. That is, there exist continuous, increasing functions $\psi_1^+, \ldots, \psi_N^+ : \mathbb{R}_+ \longrightarrow \mathbb{R}$ and $\psi_1^-, \ldots, \psi_N^- : \mathbb{R}_- \longrightarrow \mathbb{R}$ giving the additive representation (A9). Furthermore, the functions $\psi_1^+, \ldots, \psi_N^+, \psi_1^-, \ldots, \psi_N^-$ are unique up to increasing affine transformation with a common scalar multiplication.

For all $n \in [1 \ldots N]$, let $k_n^{\pm} := \psi_n^{\pm}(0)$. By replacing $\psi_n^{\pm}$ with the function $\psi_n^{\pm} - k_0^{\pm}$ if necessary, we can assume without loss of generality that $\psi_n^{\pm}(0) = 0$ for all $n \in [1 \ldots N]$. Now let $s := \psi_1^+(1)$. By replacing $\psi_n^{\pm}$ with the function $\psi_n^{\pm}/s$ for all $n \in [1 \ldots N]$ if necessary, we can assume without loss of generality that $\psi_1^+(1) = 1$.

◇ Claim 6

For all $N \in \mathbb{N}$, Claim 6 yields a collection of functions $\psi_{N,1}^+, \ldots, \psi_{N,N}^+ : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ and $\psi_{N,1}^-, \ldots, \psi_{N,N}^- : \mathbb{R}_- \longrightarrow \mathbb{R}_-$ providing an additive representation (A9) for $\succeq_N$ on $\mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow}$, and furthermore such that $\psi_{N,1}^+(1) = 1$ and $\psi_{N,n}^{\pm}(0) = 0$ for all $n \in [1 \ldots N]$.

Now, if $N < M$, then $\mathbb{R}_+^{N\downarrow}$ can be embedded into $\mathbb{R}_+^{M\downarrow}$ in a natural way, by sending $(x_1, x_2, \ldots, x_N)$ to $(x_1, x_2, \ldots, x_N, 0, 0, \ldots, 0)$ (where there are $M - N$ zeros). Likewise, $\mathbb{R}_-^{N\uparrow}$ embeds into $\mathbb{R}_-^{M\uparrow}$ in a natural way. Thus, we obtain a natural embedding of $\mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow}$ into $\mathbb{R}_+^{M\downarrow} \times \mathbb{R}_-^{M\uparrow}$. Under this embedding, the ordering $\succeq_N$ is the restriction of the ordering $\succeq_M$ to $\mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow}$ (because both arise as restrictions of the order $\succeq_*$ to their respective domains). Thus, the functions $\psi_{M,1}^+, \ldots, \psi_{M,N}^+, \psi_{M,1}^-, \ldots, \psi_{M,N}^-$ yield a *second* additive representation of $\succeq_N$. But the additive representations in Claim 6 are unique. Thus, we obtain $\psi_{M,n}^{\pm} = \psi_{N,n}^{\pm}$ for all $n \in [1 \ldots N]$. It follows that there is in fact a *single* infinite sequence of functions $(\phi_n^+)_{n=1}^{\infty}$ such that

$$\psi_{N,n}^+ = \phi_n^+, \quad \text{for all } N \in \mathbb{N} \text{ and all } n \in [1 \ldots N]. \tag{A11}$$

Likewise, there is a single infinite sequence of functions $(\phi_n^-)_{n=1}^{\infty}$ such that

$$\psi_{N,n}^- = \phi_n^-, \quad \text{for all } N \in \mathbb{N} \text{ and all } n \in [1 \ldots N]. \tag{A12}$$

It remains to show that the functions $\{\phi_n^+\}_{n=1}^{\infty}$ and $\{\phi_n^-\}_{n=1}^{\infty}$ yield the additive representation (2B) for $\succeq_*$. To see this, let $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{\infty\downarrow} \times \mathbb{R}_-^{\infty\uparrow}$. From formula (2G), there exist $L, M \in \mathbb{N}$ such that $\mathbf{x} \in \mathbb{R}_+^{L\downarrow} \times \mathbb{R}_-^{L\uparrow}$, and $\mathbf{y} \in \mathbb{R}_+^{M\downarrow} \times \mathbb{R}_-^{M\uparrow}$. Let $N := \max\{L, M\}$. Then $\mathbb{R}_+^{L\downarrow} \times \mathbb{R}_-^{L\uparrow} \subseteq \mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow}$ and $\mathbb{R}_+^{M\downarrow} \times \mathbb{R}_-^{M\uparrow} \subseteq \mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow}$. Thus, both $\mathbf{x}$ and $\mathbf{y}$ are elements of $\mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow}$, and we have

$$\begin{aligned}
\left(\mathbf{x} \succeq_* \mathbf{y}\right) \underset{\text{(a)}}{\iff} &\left(\mathbf{x} \succeq_N \mathbf{y}\right)\\
\underset{\text{(b)}}{\iff} &\left(\sum_{n=1}^{N} \psi_{N,n}^{+}(x_n^{+}) + \sum_{n=1}^{N} \psi_{N,n}^{-}(x_n^{-}) \geqslant \sum_{n=1}^{N} \psi_{N,n}^{+}(y_n^{+}) + \sum_{n=1}^{N} \psi_{N,n}^{-}(y_n^{-})\right)\\
\underset{\text{(c)}}{\iff} &\left(\sum_{n=1}^{N} \phi_n^{+}(x_n^{+}) + \sum_{n=1}^{N} \phi_n^{-}(x_n^{-}) \geqslant \sum_{n=1}^{N} \phi_n^{+}(y_n^{+}) + \sum_{n=1}^{N} \phi_n^{-}(y_n^{-})\right)\\
\underset{\text{(d)}}{\iff} &\left(\sum_{n=1}^{\infty} \phi_n^{+}(x_n^{+}) + \sum_{n=1}^{\infty} \phi_n^{-}(x_n^{-}) \geqslant \sum_{n=1}^{\infty} \phi_n^{+}(y_n^{+}) + \sum_{n=1}^{\infty} \phi_n^{-}(y_n^{-})\right),
\end{aligned}$$

as desired. Here, (a) is by the definition of $\succeq_N$, (b) is by the additive representation (A9), (c) is by Eqs. (A11) and (A12), and (d) is because $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{N\downarrow} \times \mathbb{R}_-^{N\uparrow}$, so that $x_n^+ = 0$ and $y_n^+ = 0$ for all $n \in [N+1 \ldots \infty)$. □

Remark The proof of Claim 6 uses a very similar strategy to Ebert's (1988) proof of his Theorem 1. But Ebert's proof contains an error, identified by Wakker (1993, §2.3). Fortunately, the result claimed by Ebert is actually correct (Wakker 1993, Corollary 3.6). But this result only applies to the open cone of *strictly positive* nonincreasing vectors $\mathbb{R}_{++}^{N\downarrow}$, whereas we need the corresponding result for the closed cone $\mathbb{R}_+^{N\downarrow}$ of *nonnegative* nonincreasing vectors. As shown by Wakker (1993, Example 3.8), this extension does *not* come for free; hence, the detailed argument is provided above in the proof of Claims 1–6. Despite Wakker's (1993) admonition, later authors have recapitulated Ebert's error. For example, Balasubramanian (2015, Corollary 3) repeats Ebert's proof almost verbatim. Likewise, in the proof of their Lemma 1, Asheim and Zuber (2014) cite Ebert's (1988) Theorem 1 without correction.

Proof of Proposition 2.1 (a) Let $\succeq_N$ be the restriction of the order $\succeq_*$ to $\mathbb{R}_+^{N\downarrow}$, while $\succeq_{N+1}$ is the restriction of $\succeq_*$ to $\mathbb{R}_+^{N+1\downarrow}$. Let $\mathbf{x} = (x_1, \ldots, x_N)$ and $\mathbf{y} = (y_1, \ldots, y_N)$ be in $\mathbb{R}_+^{N\downarrow}$, let $z > \max(x_1, y_1)$, and let $\mathbf{x}' := (z, x_1, \ldots, x_N)$ and $\mathbf{y}' := (z, y_1, \ldots, y_N)$. Then $\mathbf{x}', \mathbf{y}' \in \mathbb{R}_+^{N+1\downarrow}$, and we have

$$\begin{aligned}
\left(\mathbf{x} \succeq_N \mathbf{y}\right) \underset{(\dagger)}{\iff} &\left(\mathbf{x}' \succeq_{N+1} \mathbf{y}'\right)\\
\underset{(*)}{\iff} &\left(\phi_1^{+}(z) + \sum_{n=1}^{N} \phi_{n+1}^{+}(x_n) \geqslant \phi_1^{+}(z) + \sum_{n=1}^{N} \phi_{n+1}^{+}(y_n)\right)\\
\iff &\left(\sum_{n=1}^{N} \phi_{n+1}^{+}(x_n) \geqslant \sum_{n=1}^{N} \phi_{n+1}^{+}(y_n)\right).
\end{aligned}$$

Here, (†) is by Top-independence in good worlds, while $(*)$ is by the representation (2B). This equivalence holds for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{N\downarrow}$, and this argument can be repeated for any $N \in \mathbb{N}$. Thus, if we define $\psi_n^+ := \phi_{n+1}^+$ for all $n \in \mathbb{N}$, then the functions $\{\psi_n^+\}_{n=1}^\infty$ and $\{\phi_n^-\}_{n=1}^\infty$ yield *another* rank-additive representation like (2B) for $\succeq$. But the functions $\{\phi_n^\pm\}_{n=1}^\infty$ in this representation are unique up to multiplication by a common scalar. Thus, there is some $\beta > 0$ such that $\psi_n^+ = \beta\,\phi_n^+$ for all $n \in \mathbb{N}$—equivalently, $\phi_{n+1}^+ = \beta\,\phi_n^+$ for all $n \in \mathbb{N}$. Let $\phi^+ := \phi_1^+/\beta$; then, we obtain $\phi_n^+ := \beta^n\,\phi^+$ for all $n \in \mathbb{N}$. The result follows.

The proof of (b) is almost identical, but works with $\{\phi_n^-\}_{n=1}^\infty$ instead of $\{\phi_n^+\}_{n=1}^\infty$. The uniqueness claims in parts (a), (b) and (c) all follow immediately from the uniqueness statement in Theorem 1. □

Proof of Proposition 2.2 First, note that the supremum $\overline{W}$ is never obtained by any $\mathbf{x} \in \mathcal{X}$, even if $\overline{W}$ is finite. To see this, suppose by contradiction that $W(\mathbf{x}) = \overline{W}$ for some $\mathbf{x} \in \mathcal{X}$. Let $\mathbf{x}'$ be obtained by increasing $\mathbf{x}$ by some amount in every nonzero coordinate. Then $W(\mathbf{x}') > W(\mathbf{x})$, because the functions $\phi_n^\pm$ are all strictly increasing. Thus, $W(\mathbf{x}') > \overline{W}$, contradicting the definition of $\overline{W}$.

(a) "$\Longrightarrow$" Let $\mathbf{x}$ and r_0 be as in the formulation of No Repugnant Conclusion. For any $N \in \mathbb{N}$, we have $\mathbf{x} \succ r_0\mathbf{1}_N$, and thus,

$$W(\mathbf{x}) > W(r_0\mathbf{1}_N) = \sum_{n=1}^{N} \phi_n^+(r_0).$$

Taking the limit as $N\to\infty$, we obtain $\displaystyle\sum_{n=1}^{\infty} \phi_n^+(r_0) \leqslant W(\mathbf{x}) < \overline{W}$, as desired.

"$\Longleftarrow$" Let r_0 satisfy the condition in the theorem. Then there exists some $\mathbf{x} \in \mathcal{X}$ such that $W(\mathbf{x}) > \sum_{n=1}^\infty \phi_n^+(r_0)$, and thus, $W(\mathbf{x}) > \sum_{n=1}^N \phi_n^+(r_0)$ for all $N \in \mathbb{N}$. It follows that $\mathbf{x} \succ r_0\mathbf{1}_N$ for all $N \in \mathbb{N}$, as desired.

(b) "$\Longrightarrow$" For any $N \in \mathbb{N}$, let $\mathbf{x} \in \mathcal{X}$ satisfy the statement of No utility monsters. Thus, for all $r \in \mathbb{R}_+$, we have $\mathbf{x} \succ r\,\mathbf{1}_N$, and thus,

$$W(\mathbf{x}) > W(r\,\mathbf{1}_N) = \sum_{n=1}^{N} \phi_n^+(r).$$

Taking the limit as $r\to\infty$, we obtain $\displaystyle\lim_{r\to\infty}\sum_{n=1}^{N} \phi_n^+(r) \leqslant W(\mathbf{x}) < \overline{W}$, as desired.

"$\Longleftarrow$" For any $N \in \mathbb{N}$, we have $\lim_{r\to\infty}\sum_{n=1}^N \phi_n^+(r) < \overline{W}$. Thus, there exists some $\mathbf{x} \in \mathcal{X}$ such that $\lim_{r\to\infty}\sum_{n=1}^N \phi_n^+(r) < W(\mathbf{x})$. Thus, for all $r \in \mathbb{R}_+$, we have $W(r\,\mathbf{1}_N) < W(\mathbf{x})$, and thus, $r\,\mathbf{1}_N \prec \mathbf{x}$, as desired.

For the last statement of the theorem, suppose that $\overline{W} < \infty$. Let $r_0 > 0$, and let $r_1 > r_0$; then, $\sum_{n=1}^\infty \phi_n^+(r_1) \leqslant \overline{W}$. Now let $\delta := \phi_1^+(r_1) - \phi_1^+(r_0)$. Then $\delta > 0$ because ϕ_1^+ is strictly increasing, and we have

$$\sum_{n=1}^{\infty} \phi_n^+(r_1) \geqslant \delta + \sum_{n=1}^{\infty} \phi_n^+(r_0) > \sum_{n=1}^{\infty} \phi_n^+(r_0).$$

It follows that $\sum_{n=1}^{\infty} \phi_n^+(r_0) < \overline{W}$. Thus, the condition in part (a) is satisfied. (In fact, this argument works for *all* $r_0 > 0$.) By a similar argument, we deduce that $\lim_{r\to\infty} \sum_{n=1}^{N} \phi_n^+(r) < \overline{W}$, for all $N \in \mathbb{N}$. Thus, part (b) is satisfied. □

Proof of Proposition 2.3 The first statement is obvious. The second follows immediately from Proposition 2.2. □

Proof of Proposition 2.4 Before proceeding with the proof of (a), (b), and (c), we need some preliminary observations. Let $\succeq$ be a SWO on $\mathcal{X}$. Let $\succeq_*$ be the ordering on $\mathbb{R}_+^{\propto\downarrow} \times \mathbb{R}_-^{\propto\uparrow}$ defined via statement (2A).

Claim 1 $\succeq$ *satisfies* Inequality neutrality *(respectively,* Inequality aversion*, resp.* Strict inequality aversion*) on* $\mathcal{X}$ *if and only if* $\succeq_*$ *satisfies the same axiom on* $\mathbb{R}_+^{\propto\downarrow} \times \mathbb{R}_-^{\propto\uparrow}$.

Proof Let $\mathbf{x}, \mathbf{y} \in \mathcal{X}$. Say that $\mathbf{y}$ is a *rank-preserving Pigou–Dalton transform* of $\mathbf{x}$ if $\mathbf{y}$ is a Pigou–Dalton transform of $\mathbf{x}$, and furthermore, for all $i, j \in \mathcal{I}$, if $x_i < x_j$, then $y_i \leqslant y_j$; also, if $x_i < 0$, then $y_i \leqslant 0$; finally, if $x_i > 0$, then $y_i \geqslant 0$. In other words, the reallocation of utility does not change the ranking of people from best-off to worst-off which we use to apply the rank-additive SWF (2B). Note that we allow the possibility that $x_i < x_j$ but $y_i = y_j$—the reallocation may equalize two people (so that afterwards they could be ranked in either order). Likewise, we allow the possibility that $x_i < 0$ (or $x_i > 0$) but $y_i = 0$. The following facts are easily verified:

(a) For any $\mathbf{x}, \mathbf{z} \in \mathcal{X}$, $\mathbf{z}$ is an ordinary Pigou–Dalton transform of $\mathbf{x}$ if and only if there is a sequence $\mathbf{x} = \mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N = \mathbf{z}$ such that for all $n \in [1 \ldots N]$, $\mathbf{y}_n$ is a rank-preserving Pigou–Dalton transform of $\mathbf{y}_{n-1}$.
(b) For any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, if $\mathbf{y}$ is a rank-preserving Pigou–Dalton transform of $\mathbf{x}$, then $(\mathbf{y}^+, \mathbf{y}^-)$ is an ordinary Pigou–Dalton transform of $(\mathbf{x}^+, \mathbf{x}^-)$.

Fact (a) means that $\succeq$ satisfies Inequality neutrality (resp. Inequality aversion, resp. Strict inequality aversion) with respect to rank-preserving Pigou–Dalton transforms if and only if it satisfies this axiom with respect to *all* Pigou–Dalton transforms. Fact (b) means that $\succeq$ satisfies one of these three axioms with respect to rank-preserving Pigou–Dalton transforms if and only if $\succeq_*$ satisfies the corresponding axiom (in its ordinary form) on $\mathbb{R}_+^{\propto\downarrow} \times \mathbb{R}_-^{\propto\uparrow}$. This proves the claim. ◇ Claim 1

Now let $\mathbf{x} = (\mathbf{x}^+, \mathbf{y}^-)$ and $\mathbf{y} = (\mathbf{y}^+, \mathbf{y}^-)$ be elements of $\mathbb{R}_+^{\propto\downarrow} \times \mathbb{R}_-^{\propto\uparrow}$, and suppose $\mathbf{y}$ is a Pigou–Dalton transform of $\mathbf{x}$. Then there exist $m, n \in \mathbb{N}$ and $\epsilon > 0$ such that one of the following three cases occurs:

(i) $y_m^- = x_m^- + \epsilon \leqslant 0 \leqslant y_n^+ = x_n^+ - \epsilon$, while $y_\ell^- = x_\ell^-$ for all $\ell \in \mathbb{N} \setminus \{m\}$, and $y_\ell^+ = x_\ell^+$ for all $\ell \in \mathbb{N} \setminus \{n\}$.
(ii) $m > n$, and $y_m^+ = x_m^+ + \epsilon \leqslant y_n^+ = x_n^+ - \epsilon$, while $y_\ell^+ = x_\ell^+$ for all $\ell \in \mathbb{N} \setminus \{m, n\}$, and $y_\ell^- = x_\ell^-$ for all $\ell \in \mathbb{N}$.

(iii) $m < n$, and $y_m^- = x_m^- + \epsilon \leqslant y_n^- = x_n^- - \epsilon$, while $y_\ell^- = x_\ell^-$ for all $\ell \in \mathbb{N} \setminus \{m, n\}$, and $y_\ell^+ = x_\ell^+$ for all $\ell \in \mathbb{N}$.

Let W be the SWF in formula (2B). The $W(\mathbf{y}) - W(\mathbf{x})$ takes the following form in Cases (i), (ii), and (iii):

$$\begin{aligned}
&\text{(I)} \quad W(\mathbf{y}) - W(\mathbf{x}) = \left[\phi_m^-(x_m^- + \epsilon) - \phi_m^-(x_m^-)\right] - \left[\phi_n^+(x_n^+) - \phi_n^+(x_n^+ - \epsilon)\right].\\
&\text{(II)} \quad W(\mathbf{y}) - W(\mathbf{x}) = \left[\phi_m^+(x_m^+ + \epsilon) - \phi_m^+(x_m^+)\right] - \left[\phi_n^+(x_n^+) - \phi_n^+(x_n^+ - \epsilon)\right].\\
&\text{(III)} \quad W(\mathbf{y}) - W(\mathbf{x}) = \left[\phi_m^-(x_m^- + \epsilon) - \phi_m^-(x_m^-)\right] - \left[\phi_n^-(x_n^-) - \phi_n^-(x_n^- - \epsilon)\right].
\end{aligned}$$

With these preliminaries established, we proceed with the proof of parts (a), (b), and (c) of the theorem. In each of (a), (b), and (c), it is easily verified that the stated conditions are sufficient for $\succeq_*$ to satisfy the stated axiom—and hence, for $\succeq$ to satisfy it, by Claim 1. It remains to prove that they are also necessary.

(a) Suppose $\succeq$ (and hence, $\succeq_*$) satisfies Inequality neutrality. So if $\mathbf{y}$ is a Pigou–Dalton transform of $\mathbf{x}$, then $W(\mathbf{y}) = W(\mathbf{x})$. Thus, for any $m, n \in \mathbb{N}$, any $\epsilon > 0$, and any $x_m^- < -\epsilon$ and $x_n^+ > \epsilon$, the right-hand side of equation (I) is zero. Thus, there is some constant $C > 0$ such that $\phi_m^-(x_m^- + \epsilon) - \phi_m^-(x_m^-) = C$ and $\phi_n^+(x_n^+) - \phi_n^+(x_n^+ - \epsilon) = C$ for all $x_m^- < -\epsilon$ and $x_n^+ > \epsilon$. Thus, ϕ_n^+ and ϕ_m^- must each have a constant slope—in fact, the *same* slope. Since $\phi_n^+(0) = 0$ and $\phi_m^-(0) = 0$ by assumption, this means they are linear functions with the same slope. Varying this argument over all $m, n \in \mathbb{N}$, we conclude that the $\{\phi_n^+\}_{n=1}^\infty$ and $\{\phi_n^-\}_{n=1}^\infty$ are all linear functions with the same slope. Thus, SWF (2B) is equivalent (up to multiplication by a scalar) to the classical utilitarian SWF (2C).

(b) Suppose $\succeq$ (and hence, $\succeq_*$) satisfies Inequality aversion. So if $\mathbf{y}$ is a Pigou–Dalton transform of $\mathbf{x}$, then $W(\mathbf{y}) \geqslant W(\mathbf{x})$. Thus, for any $m, n \in \mathbb{N}$, any $\epsilon > 0$, and any $x_n^\pm, x_m^\pm \in \mathbb{R}$, we have:

- If $x_m^- < -\epsilon$ and $x_n^+ > \epsilon$, then the right-hand side of equation (I) is nonnegative.
- If $x_n^+ - 2\epsilon \geqslant x_m^+ \geqslant 0$, then the right-hand side of equation (II) is nonnegative.
- If $0 \geqslant x_n^- \geqslant x_m^- + 2\epsilon$, then the right-hand side of equation (III) is nonnegative.

Setting $s := x_m^\pm + \epsilon$ and $r := x_n^\pm - \epsilon$ in all three cases, we obtain inequalities (i), (ii), and (iii) in part (b) of the theorem.

To obtain inequality (2H), let $J \in \mathbb{N}$, and let $\epsilon := q/J$. Then for any $n < m \in \mathbb{N}$,

$$\begin{aligned}
\phi_n^+(q) = \phi_n^+(q) - \phi_n^+(0) &= \sum_{j=0}^{J-1} \left(\phi_n^+\left((j+1)\epsilon\right) - \phi_n^+\left(j\epsilon\right)\right)\\
&= \left(\phi_n^+(\epsilon) - \phi_n^+(0)\right) \; + \; \sum_{j=1}^{J} \left(\phi_n^+\left((j+1)\epsilon\right) - \phi_n^+\left(j\epsilon\right)\right)\\
&\qquad\qquad - \left(\phi_n^+\left((J+1)\epsilon\right) - \phi_n^+\left(J\epsilon\right)\right)\\
&\underset{(*)}{\leqslant} \left(\phi_n^+(\epsilon) - \phi_n^+(0)\right) \; + \; \sum_{j=1}^{J} \left(\phi_m^+\left(j\epsilon\right) - \phi_m^+\left((j-1)\epsilon\right)\right)
\end{aligned}$$

$$\begin{aligned} &\qquad - \left(\phi_n^+(q+\epsilon) - \phi_n^+(q)\right) \\ &= \phi_n^+\left(\frac{q}{J}\right) \ + \ \phi_m^+(q) \ - \ \left(\phi_n^+\left(q+\frac{q}{J}\right) - \phi_n^+(q)\right). \end{aligned}$$

Here, $(*)$ is by inequality (b)(ii), where for each summand, we set $r = s = j\epsilon$, so that $r+\epsilon = (j+1)\epsilon$ and $s-\epsilon = (j-1)\epsilon$. We have also used several times the fact that $\phi_n^+(0) = \phi_m^+(0) = 0$. Taking the limit as $J \to \infty$, we obtain:

$$\begin{aligned} \phi_n^+(q) \leqslant \phi_m^+(q) \ + \ &\lim_{J\to\infty} \phi_n^+\left(\frac{q}{J}\right) \\ &- \lim_{J\to\infty} \left(\phi_n^+\left(q+\frac{q}{J}\right) - \phi_n^+(q)\right) = \phi_m^+(q), \end{aligned}$$

where the last step is because ϕ_n^+ is continuos at 0 and at q. Thus, we deduce that $\phi_n^+(q) \leqslant \phi_m^+(q)$ for all $q \in \mathbb{R}_+$ and $n < m \in \mathbb{N}$. This justifies all the inequalities on the left side of (2H). By an almost identical argument [using inequality (b)(iii)], we deduce that $\phi_n^-(q) \leqslant \phi_m^-(q)$ for all $q \in \mathbb{R}_+$ and $n > m \in \mathbb{N}$; this justifies all the inequalities on the right side of (2H). Finally, by a similar argument [using inequality (b)(i)], we deduce that $\phi_n^+(q) \leqslant \phi_m^-(q)$ for all $q \in \mathbb{R}_+$ and all $n, m \in \mathbb{N}$. This justifies the inequalities between the left and right sides of (2H).

To prove inequality (2I), observe that inequalities (b)(i)–(b)(iii) imply the following

(i) If $r \geqslant 0 \geqslant s$, then $\dfrac{\phi_n^+(r+\epsilon) - \phi_n^+(r)}{\epsilon} \leqslant \dfrac{\phi_m^-(s) - \phi_m^-(s-\epsilon)}{\epsilon}$.

(ii) If $n < m$ and $r \geqslant s \geqslant \epsilon > 0$, then $\dfrac{\phi_n^+(r+\epsilon) - \phi_n^+(r)}{\epsilon} \leqslant \dfrac{\phi_m^+(s) - \phi_m^+(s-\epsilon)}{\epsilon}$.

(iii) If $n > m$ and $s \leqslant r \leqslant -\epsilon < 0$, then $\dfrac{\phi_n^-(r+\epsilon) - \phi_n^-(r)}{\epsilon} \leqslant \dfrac{\phi_m^-(s) - \phi_m^-(s-\epsilon)}{\epsilon}$.

Taking the limit as $\epsilon \to 0$ in all three cases, we deduce:

(i′) If $r \geqslant 0 \geqslant s$, then $(\phi_n^+)'(r) \leqslant (\phi_m^-)'(s)$.
(ii′) If $n < m$ and $r \geqslant s > 0$, then $(\phi_n^+)'(r) \leqslant (\phi_m^+)'(s)$.
(iii′) If $n > m$ and $s \leqslant r < 0$, then $(\phi_n^-)'(r) \leqslant (\phi_m^-)'(s)$.

If $r_1 \geqslant r_2 \geqslant r_3 \geqslant \cdots \geqslant 0$ and $s_1 \leqslant s_2 \leqslant s_3 \leqslant \cdots \leqslant 0$, then each of the inequalities in between adjacent terms in (2I) can be obtained by invoking one of inequalities (i′), (ii′), or (iii′).

(c) The proof is identical to (b), but with strict inequalities.

□

Proof of Proposition 2.6 Easy modification of the proof of Proposition 2.4. □

B Appendix: Proofs from Section 3

Parts of the proof of Theorem 2 are analogous to parts of the proof of Theorem 1. When noting these analogies, I will refer to Claim N in the proof of Theorem 1 as "Claim 1.N".

Proof of Theorem 2 The proof of "$\Longleftarrow$" is straightforward, so I will focus on the proof of "$\Longrightarrow$". First, Claims 1–6 will show that each of the orders $\succeq_N$ admits an additive representation on $\mathbb{R}^{N\uparrow}$. Then, Claims 7–14 will combine all these representations together to obtain an ARA SWF on $\mathbb{R}^{\propto\uparrow}$.

The binary relation "$\cong$" in the Trade-off Consistency axiom is clearly symmetric: if $(a \overset{n}{\rightsquigarrow} b) \cong (c \overset{m}{\rightsquigarrow} d)$, then $(c \overset{m}{\rightsquigarrow} d) \cong (a \overset{n}{\rightsquigarrow} b)$. Claim 1 shows that $\cong$ is also "transitive".

Claim 1 *Let $n, m, \ell \in \mathbb{N}$ be distinct, and let $a, b, c, d, e, f \in \mathbb{R}$. If $(a \overset{n}{\rightsquigarrow} b) \cong (c \overset{m}{\rightsquigarrow} d)$ and $(c \overset{m}{\rightsquigarrow} d) \cong (e \overset{\ell}{\rightsquigarrow} f)$, then $(a \overset{n}{\rightsquigarrow} b) \cong (e \overset{\ell}{\rightsquigarrow} f)$.*[28]

Proof For any $\mathbf{x} \in \mathcal{X}$, if $x_n^{\uparrow} = a$ and $x_m^{\uparrow} = c$, and $b_{(n)}\mathbf{x}$ and $d_{(m)}\mathbf{x}$ are well defined,[29] then Trade-off Consistency says that

$$b_{(n)}\mathbf{x} \approx d_{(m)}\mathbf{x}, \tag{B1}$$

because $(a \overset{n}{\rightsquigarrow} b) \cong (c \overset{m}{\rightsquigarrow} d)$. Likewise, for any $\mathbf{x} \in \mathcal{X}$, if $x_m^{\uparrow} = c$ and $x_\ell^{\uparrow} = e$, and $d_{(m)}\mathbf{x}$ and $f_{(\ell)}\mathbf{x}$ are well defined, then Trade-off Consistency says that

$$d_{(m)}\mathbf{x} \approx f_{(\ell)}\mathbf{x}, \tag{B2}$$

because $(c \overset{m}{\rightsquigarrow} d) \cong (e \overset{\ell}{\rightsquigarrow} f)$.

Now, find $\mathbf{x} \in \mathcal{X}$ such that $x_n^{\uparrow} = a$, $x_m^{\uparrow} = c$, and $x_\ell^{\uparrow} = e$, and such that $b_{(n)}\mathbf{x}$, $d_{(m)}\mathbf{x}$, and $f_{(\ell)}\mathbf{x}$ are all well defined. Then combining (B1) and (B2) and the transitivity of the indifference relation $\approx$, we get $b_{(n)}\mathbf{x} \approx f_{(\ell)}\mathbf{x}$. Thus, $(a \overset{n}{\rightsquigarrow} b) \cong (e \overset{\ell}{\rightsquigarrow} f)$.

◇ Claim 1

Let $\mathbb{R}^{N\uparrow\uparrow} := \{\mathbf{x} \in \mathbb{R}^N;\ x_1 < x_2 < \cdots < x_N\}$; this is the topological interior of $\mathbb{R}^{N\uparrow}$ as a subset of $\mathbb{R}^N$. Let $\mathcal{W} \subseteq \mathbb{R}^{N\uparrow\uparrow}$. The order $\succeq_N$ is *coordinate-independent* on $\mathcal{W}$ if the following is true: for any $\mathbf{w}, \mathbf{v}, \mathbf{w}', \mathbf{v}' \in \mathbb{R}^{N\uparrow}$, and any $n \in [1 \ldots N]$ such that $w_n = v_n$ and $w'_n = v'_n$, while $\mathbf{w}_{-n} = \mathbf{w}'_{-n}$ and $\mathbf{v}_{-n} = \mathbf{v}'_{-n}$, we have $\mathbf{w} \succeq_N \mathbf{v}$ if and only if $\mathbf{w}' \succeq_N \mathbf{v}'$. Say that $\succeq_N$ is *locally coordinate-independent* on $\mathbb{R}^{N\uparrow\uparrow}$ if it is coordinate-independent in an open neighbourhood of every point in $\mathbb{R}^{N\uparrow\uparrow}$.

Claim 2 *For every $N \in \mathbb{N}$, the order $\succeq_N$ is locally coordinate-independent on $\mathbb{R}^{N\uparrow\uparrow}$.*

Proof Let $\mathbf{x} \in \mathbb{R}^{N\uparrow\uparrow}$. Since $\mathbb{R}^{N\uparrow\uparrow}$ is an open subset of $\mathbb{R}^N$, it contains an open neighbourhood around $\mathbf{x}$—in other words, $\mathbf{y} \in \mathbb{R}^{N\uparrow\uparrow}$ for all points $\mathbf{y} \in \mathbb{R}^N$ that are "close enough" to $\mathbf{x}$. Throughout this proof, when I use the words *close enough*, I will mean them in this way: the construction I want to perform involves a small enough change in coordinate values that it does not change the strict ordering of the coordinates.

Now, fix $n \in [1 \ldots N]$. Let $\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}'$ all be close enough together, and suppose that $x_n = y_n$ and $x'_n = y'_n$, while $\mathbf{x}_{-n} = \mathbf{x}'_{-n}$ and $\mathbf{y}_{-n} = \mathbf{y}'_{-n}$. I will show that $\mathbf{x} \succeq_N \mathbf{y}$ if and only if $\mathbf{x}' \succeq_N \mathbf{y}'$.

[28] Here, I assume that the ordering of a, b versus c, d versus e, f is the same as the ordering of n, m and ℓ. For example, if $n < m < \ell$, then $a, b < c, d < e, f$.

[29] That is: $x_{n-1}^{\uparrow} \leqslant b \leqslant x_{n+1}^{\uparrow}$ and $x_{m-1}^{\uparrow} \leqslant d \leqslant x_{m+1}^{\uparrow}$.

For simplicity, suppose $n = 1$ (the same argument works in general). So $\mathbf{x} = (a, x_2, x_3, \ldots, x_N)$ and $\mathbf{y} = (a, y_2, y_3, \ldots, y_N)$, while $\mathbf{x}' = (b, x_2, x_3, \ldots, x_N)$ and $\mathbf{y}' = (b, y_2, y_3, \ldots, y_N)$, for some $x_2 < \cdots < x_N$ and $y_2 < \cdots < y_N$ and some $a, b < \min\{x_2, y_2\}$.

Let $a^1 := a$. If y_2 is close enough to x_2, then Continuity and Pareto yield some $a^2 \in \mathbb{R}$ such that $(a^2, x_2, x_3, x_4, \ldots, x_N) \approx_N (a^1, y_2, x_3, x_4, \ldots, x_N)$.[30] In other words, $(a^1 \overset{1}{\rightsquigarrow} a^2) \cong (x_2 \overset{2}{\rightsquigarrow} y_2)$.

Next, if y_3 is close enough to x_3, then Continuity and Pareto yield some $a^3 \in \mathbb{R}$ such that $(a^3, x_2, x_3, x_4, \ldots, x_N) \approx_N (a^2, x_2, y_3, x_4, \ldots, x_N)$. In other words, $(a^2 \overset{1}{\rightsquigarrow} a^3) \cong (x_3 \overset{3}{\rightsquigarrow} y_3)$.

Now let $n \in [4 \ldots N]$, and suppose we have constructed a^{n-1}. If y_n is close enough to x_n, then Continuity and Pareto yield some $a^n \in \mathbb{R}$ such that

$$(a^n, x_2, \ldots, x_{n-1}, x_n, x_{n+1}, \ldots, x_N) \approx_N (a^{n-1}, x_2, \ldots, x_{n-1}, y_n, x_{n+1}, \ldots, x_N).$$

In other words,

$$(a^{n-1} \overset{1}{\rightsquigarrow} a^n) \cong (x_n \overset{n}{\rightsquigarrow} y_n), \quad \text{for all } n \in [2 \ldots N]. \tag{B3}$$

Now, by repeatedly applying Trade-off Consistency to the relations (B3), we obtain:

$$\begin{aligned} \mathbf{y} = (a^1, y_2, y_3, y_4, \ldots, y_N) &\approx_N (a^2, x_2, y_3, y_4, \ldots, y_N) \\ &\approx_N (a^3, x_2, x_3, y_4, \ldots, y_N) \approx_N \cdots \\ \cdots &\approx_N (a^N, x_2, x_3, x_4, \ldots, x_N). \end{aligned} \tag{B4}$$

(If $\mathbf{x}$ and $\mathbf{y}$ are close enough, then all of these intermediate vectors are in $\mathbb{R}^{N}_{\uparrow\uparrow}$.) Thus,

$$\begin{aligned} \left(\mathbf{x} \succeq_N \mathbf{y}\right) &\underset{(*)}{\iff} \left((a^1, x_2, x_3, \ldots, x_N) \succeq_N (a^N, x_2, x_3, \ldots, x_N)\right) \\ &\underset{(\dagger)}{\iff} \left(a^1 \geqslant a^N\right). \end{aligned} \tag{B5}$$

Here $(*)$ is because $\mathbf{x} = (a^1, x_2, x_3, \ldots, x_N)$ (because $a^1 = a = x_1$) and $\mathbf{y} \approx_N (a^N, x_2, x_3, \ldots, x_N)$ by formula (B4). Meanwhile (†) is by Pareto.

Now let $b^1 := b$. If $\mathbf{x}$ and $\mathbf{y}$ are close enough, then by repeating the preceding argument, we can construct a sequence $b^2, b^3, \ldots, b^N \in \mathbb{R}$ such that

$$(b^{n-1} \overset{1}{\rightsquigarrow} b^n) \cong (x_n \overset{n}{\rightsquigarrow} y_n), \quad \text{for all } n \in [2 \ldots N]. \tag{B6}$$

By repeatedly applying Trade-off Consistency to (B6), we obtain:

$$\begin{aligned} \mathbf{y}' = (b^1, y_2, y_3, y_4, \ldots, y_N) &\approx_N (b^2, x_2, y_3, y_4, \ldots, y_N) \\ &\approx_N (b^3, x_2, x_3, y_4, \ldots, y_N) \approx_N \cdots \\ \cdots &\approx_N (b^N, x_2, x_3, x_4, \ldots, x_N), \end{aligned}$$

[30] See, for example, the proofs of Claims 1.2 and 1.4, or Claim 4, for similar constructions.

where, if $\mathbf{x}'$ and $\mathbf{y}'$ are close enough, then all of these vectors are in $\mathbb{R}^{N}_{\uparrow\uparrow}$. Thus,

$$\begin{aligned}\left(\mathbf{x}' \succeq_N \mathbf{y}'\right) &\iff \left((b^1, x_2, x_3, \ldots, x_N) \succeq_N (b^N, x_2, x_3, \ldots, x_N)\right)\\ &\iff \left(b^1 \geqslant b^N\right). \end{aligned} \tag{B7}$$

Now I will show that $a^1 \geqslant a^N$ if and only if $b^1 \geqslant b^N$. Let $\mathbf{z} = (z_1, z_2, \ldots, z_{N+1})$ be some element of $\mathbb{R}^{N+1}_{\uparrow\uparrow}$ such that $z_1 = a^1$ and $z_2 > \max\{a_2, a_3, \ldots, a_N, b_2, b_3, \ldots, b_N\}$. Let $c^1 := z_{N+1}$. Using Pareto and Continuity, we can construct $c^2, c^3, \ldots, c^N$ such that

$$\begin{aligned}(a^1, z_2, \ldots, z_N, c^1) &\approx_{N+1} (a^2, z_2, \ldots, z_N, c^2)\\ &\approx_{N+1} (a^3, z_2, \ldots, z_N, c^3) \approx_{N+1} \cdots\\ \cdots &\approx_{N+1} (a^N, z_2, \ldots, z_N, c^N),\end{aligned} \tag{B8}$$

and all these vectors are in $\mathbb{R}^{N+1}_{\uparrow}$. (This is possible if z_{N+1} is large enough, and $|a_n - a_{n-1}|$ is small enough for all $n \in [2 \ldots N]$, which in turn is the case as long as $\mathbf{y}$ is close enough to $\mathbf{x}$.) In other words,

$$(c^n \overset{N+1}{\rightsquigarrow} c^{n-1}) \cong (a^{n-1} \overset{1}{\rightsquigarrow} a^n), \quad \text{for all } n \in [2 \ldots N]. \tag{B9}$$

Since $1 \neq n \neq N+1$, we can combine equations (B3) and (B9) via Claim 1, to get

$$(c^n \overset{N+1}{\rightsquigarrow} c^{n-1}) \cong (x_n \overset{n}{\rightsquigarrow} y_n), \quad \text{for all } n \in [2 \ldots N]. \tag{B10}$$

Then, combining equations (B6) and (B10) via Claim 1, we get

$$(c^n \overset{N+1}{\rightsquigarrow} c^{n-1}) \cong (b^{n-1} \overset{1}{\rightsquigarrow} b^n), \quad \text{for all } n \in [2 \ldots N]. \tag{B11}$$

Thus, Trade-off Consistency yields

$$\begin{aligned}(b^1, z_2, \ldots, z_N, c^1) &\approx_{N+1} (b^2, z_2, \ldots, z_N, c^2)\\ &\approx_{N+1} (b^3, z_2, \ldots, z_N, c^3) \approx_{N+1} \cdots\\ \cdots &\approx_{N+1} (b^N, z_2, \ldots, z_N, c^N).\end{aligned} \tag{B12}$$

Thus,

$$\left(a^1 \geqslant a^N\right) \underset{(*)}{\iff} \left(c^1 \leqslant c^N\right) \underset{(\dagger)}{\iff} \left(b^1 \geqslant b^N\right), \tag{B13}$$

where $(*)$ is by (B8), $(\dagger)$ is by (B12), and both use Pareto. Putting it all together, we obtain

$$\left(\mathbf{x} \succeq_N \mathbf{y}\right) \underset{(*)}{\iff} \left(a^1 \geqslant a^N\right) \underset{(\dagger)}{\iff} \left(b^1 \geqslant b^N\right) \underset{(\ddagger)}{\iff} \left(\mathbf{x}' \succeq_N \mathbf{y}'\right), \tag{B14}$$

as desired. Here, $(*)$ is by statement (B5), $(\dagger)$ is by statement (B13), and $(\ddagger)$ is by statement (B7).

For any $\mathbf{x} \in \mathbb{R}^{N}_{\uparrow\uparrow}$, we can obtain the equivalence (B14) for all $\mathbf{y}, \mathbf{x}', \mathbf{y}' \in \mathbb{R}^{N}_{\uparrow\uparrow}$ that are close enough to $\mathbf{x}$. A similar argument works for all $n \in [1 \ldots N]$. Thus, $\succeq_N$ is locally coordinate-independent on $\mathbb{R}^{N}_{\uparrow\uparrow}$. ◇ Claim 2

Using Claim 2 and a result of Wakker (1988), I will soon show that $\succeq_N$ admits a "local" additive representation in a neighbourhood of each point in $\mathbb{R}^{N}_{\uparrow\uparrow}$. I will then combine all these local additive representations using a theorem of Chateauneuf and Wakker (1993). Just as in the proof of Theorem 1, I must first check that all the technical conditions of the Chateauneuf–Wakker theorem are satisfied; this is the purpose of Claims 3 to 5.

Claim 3 *Let $N \in \mathbb{N}$. Every indifference set of $\succeq_N$ in $\mathbb{R}^{N}_{\uparrow\uparrow}$ is path-connected.*

The proof of Claim 3 is similar to the proof of Claim 1.1, but somewhat simpler. First we need a preliminary result. Let $\mathcal{Y}^N := \{\mathbf{x} \in \mathbb{R}^{N}_{\uparrow\uparrow};\ x_1 = 0\}$. For any $\mathbf{y} = (y_1, \ldots, y_N) \in \mathbb{R}^{N}_{\uparrow\uparrow}$ and $r \in \mathbb{R}$, define $\tau^r(\mathbf{y}) := (y_1 + r, y_2 + r, \ldots, y_N + r)$. The next claim is analogous to Claim 1.2.

Claim 4 *Let $N \in \mathbb{N}$, and let $\mathbf{x} \in \mathbb{R}^{N}_{\uparrow\uparrow}$. Let $\mathcal{Z} := \{\mathbf{z} \in \mathbb{R}^{N}_{\uparrow\uparrow};\ \mathbf{z} \approx_N \mathbf{x}\}$ be the indifference set of $\mathbf{x}$. For any $\mathbf{y} \in \mathcal{Y}^N$, there is a unique $r \in \mathbb{R}$ with $\tau^r(\mathbf{y}) \in \mathcal{Z}$. Let $\phi(\mathbf{y}) := \tau^r(\mathbf{y})$; this defines a continuous surjection $\phi : \mathcal{Y}^N \longrightarrow \mathcal{Z}$.*

Proof *Existence and uniqueness* Since $\mathbb{R}^{N}_{\uparrow\uparrow}$ is a connected, separable topological space, and $\succeq_N$ satisfies Continuity, the theorem of Debreu (1954) yields a continuous function $w : \mathbb{R}^{N}_{\uparrow\uparrow} \longrightarrow \mathbb{R}$ that represents $\succeq_N$. If r is large enough, then every coordinate of $\tau^r(\mathbf{y})$ is bigger than every coordinate of $\mathbf{x}$, so Pareto says that $\tau^r(\mathbf{y}) \succeq_N \mathbf{x}$, and hence, $w[\tau^r(\mathbf{y})] \geqslant w(\mathbf{x})$. If r is small enough, then every coordinate of $\tau^r(\mathbf{y})$ is smaller than every coordinate of $\mathbf{x}$, so Pareto says that $\tau^r(\mathbf{y}) \preceq_N \mathbf{x}$, and hence $w[\tau^r(\mathbf{y})] \leqslant w(\mathbf{x})$. The function $w \mapsto w[\tau^r(\mathbf{y})]$ is continuous, so the intermediate value theorem yields some $r \in \mathbb{R}$ such that $w[\tau^r(\mathbf{y})] = w(\mathbf{x})$, and hence, $\tau^r(\mathbf{y}) \approx_N \mathbf{x}$. Thus, $\tau^r(\mathbf{y}) \in \mathcal{Z}$. This value of r is unique by Pareto.

Surjective Let $\mathbf{z} \in \mathcal{Z}$. Let $\mathbf{y} := \tau^{-z_1}(\mathbf{z})$. Then $y_1 = 0$ by construction, so $\mathbf{y} \in \mathcal{Y}^N$. Clearly, $\tau^{z_1}(\mathbf{y}) = \mathbf{z}$. Thus, $\phi(\mathbf{y}) = \mathbf{z}$.

Continuous In fact, ϕ is *Lipschitz* continuous. To see this, let $\mathbf{u}, \mathbf{v} \in \mathcal{Y}^N$ and let $\epsilon := \|\mathbf{u} - \mathbf{v}\|$. Then $|u_n - v_n| \leqslant \epsilon$ for all $n \in [1 \ldots N]$. Let $\mathbf{u}' := \phi(\mathbf{u})$ and $\mathbf{v}' := \phi(\mathbf{v})$; then, there exist $r, s \in \mathbb{R}$ such that $u'_n = u_n + r$ and $v'_n = v_n + s$ for all $n \in [1 \ldots N]$.

Suppose $r > s + \epsilon$. Then for all $n \in [1 \ldots N]$, we have $u'_n = u_n + r > u_n + s + \epsilon \geqslant v_n + s = v'_n$. So $\mathbf{u}' \succ_N \mathbf{v}$ by Pareto. This contradicts the fact that both $\mathbf{u}'$ and $\mathbf{v}'$ are in the same indifference set $\mathcal{Z}$.

Likewise, if $r < s - \epsilon$, then $u'_n < v'_n$ for all $n \in [1 \ldots N]$, so $\mathbf{u}' \prec_N \mathbf{v}$ by Pareto, again contradicting the fact that they are in the same indifference set. To avoid these contradictions, we must have $|r - s| \leqslant \epsilon$. But $\|\mathbf{u}' - \mathbf{v}'\| \leqslant \|\mathbf{u} - \mathbf{v}\| + N|r - s|$, and $\epsilon = \|\mathbf{u} - \mathbf{v}\|$. Thus, $\|\mathbf{u}' - \mathbf{v}'\| \leqslant (N + 1)\, \|\mathbf{u} - \mathbf{v}\|$.

For any $\mathbf{u}, \mathbf{v} \in \mathcal{Y}^N$, this argument yields $\|\phi(\mathbf{u}) - \phi(\mathbf{v})\| \leqslant (N + 1)\|\mathbf{u} - \mathbf{v}\|$.
◇ Claim 4

***Proof of Claim* 3** $\mathcal{Y}^N$ is the intersection of the hyperplane $\{\mathbf{x} \in \mathbb{R}^N;\ x_1 = 0\}$ with the convex set $\mathbb{R}^{N\uparrow\uparrow}$, so $\mathcal{Y}^N$ is convex, hence path-connected. If $\mathcal{Z}$ is any indifference set of $\succeq_N$, then Claim 4 says that $\mathcal{Z}$ is the image of $\mathcal{Y}^N$ under a continuous surjection; hence, $\mathcal{Z}$ is also path-connected. ◇ Claim 3

Recall the *matching* terminology introduced by Chateauneuf and Wakker (1993) and reviewed in the proof of Theorem 1. The next result plays the role of Claim 1.5.

Claim 5 *Every element of* $\mathbb{R}^{N\uparrow}$ *is matched.*

Proof Let $\mathbf{x} \in \mathbb{R}^{N\uparrow}$. It is easy to see that is some $\mathbf{y} \in \mathbb{R}^{N\uparrow\uparrow}$ such that $\mathbf{x} \approx_N \mathbf{y}$. (This can be proved by a perturbation argument using Pareto and Continuity. It is similar to the proof of Claim 1.4, so the details are left to the reader.) Second, for every $n \in [1 \ldots N]$, we can find some $\mathbf{y} \in \mathbb{R}^{N\uparrow\uparrow}$ such that $y_n = x_n$. Thus, $\mathbf{x}$ is interior-matched, and hence, matched. ◇ Claim 5

The next result is analogous to Claim 1.6.

Claim 6 *For any* $N \in \mathbb{N}$ *with* $N \geqslant 3$, *there exists a unique collection of functions* $\psi_1^N, \ldots, \psi_N^N : \mathbb{R} \longrightarrow \mathbb{R}$ *with* $\psi_1^N(1) = 1$ *and with* $\psi_n^N(0) = 0$ *for all* $n \in [1 \ldots N]$, *such that for any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{N\uparrow}$, *we have*

$$\left(\mathbf{x} \succeq_N \mathbf{y}\right) \iff \left(\sum_{n=1}^{N} \psi_n^N(x_n) \geqslant \sum_{n=1}^{N} \psi_n^N(y_n)\right). \tag{B15}$$

Proof The proof strategy is similar to the proof of Claim 1.6: first, I construct a "local" additive representation in a neighbourhood of each point, and then, I stitch these local representations together using a result of Chateauneuf and Wakker (1993).

For any $\mathbf{x} \in \mathbb{R}^{N\uparrow\uparrow}$, Claim 2 yields an open neighbourhood $\mathcal{O} \subset \mathbb{R}^{N\uparrow\uparrow}$ with $\mathbf{x} \in \mathcal{O}$, such that $\succeq_N$ is coordinate-independent when restricted to $\mathcal{O}$. Let $\mathcal{B} \subset \mathcal{O}$ be an open box containing $\mathbf{x}$—say, $\mathcal{B} = (a_1, z_1) \times \cdots \times (a_N, z_N)$ for some $a_1 < z_1, \ldots, a_N < z_N$. Let $\succeq_{\mathcal{B}}$ be the restriction of $\succeq_N$ to $\mathcal{B}$. Then $\succeq_{\mathcal{B}}$ is coordinate-independent (by Claim 2) and continuous (by Continuity). Thus, Theorem 4.1 of Wakker (1988) says that there are continuous, increasing functions $\psi_n^{\mathcal{B}} : (a_n, z_n) \longrightarrow \mathbb{R}$ for all $n \in [1 \ldots N]$ such that, for any $\mathbf{b}, \mathbf{c} \in \mathcal{B}$, we have

$$\left(\mathbf{b} \succeq_{\mathcal{B}} \mathbf{c}\right) \iff \left(\sum_{n=1}^{N} \psi_n^{\mathcal{B}}(b_n) \geqslant \sum_{n=1}^{N} \psi_n^{\mathcal{B}}(c_n)\right). \tag{B16}$$

$\mathbb{R}^{N\uparrow\uparrow}$ is open, so it can be covered by such open boxes. Thus, $\succeq_N$ admits "local" additive representations (B16) everywhere on $\mathbb{R}^{N\uparrow\uparrow}$. Since $\mathbb{R}^{N\uparrow\uparrow}$ is convex, it clearly satisfies conditions (1) and (2) in Structural Assumption 2.1 of Chateauneuf and Wakker (1993). Meanwhile, condition (3) of Chateauneuf and Wakker (1993) is true by Claim 3. Finally, $\mathbb{R}^{N\uparrow\uparrow}$ is the interior of $\mathbb{R}^{N\uparrow}$, and every coordinate of any element on the boundary of $\mathbb{R}^{N\uparrow}$ is matched, by Claim 5. Thus, by Theorem 3.3 of Chateauneuf and Wakker (1993), the local additive representations (B16) can be combined together to yield a single *global* additive representation (B15) on all of $\mathbb{R}^{N}_{\uparrow}$. Furthermore, the functions $\psi_1, \ldots, \psi_N$ are unique up to increasing affine transformation with a common scalar multiplication.

For all $n \in [1 \ldots N]$, let $k_n := \psi_n(0)$. By replacing ψ_n with the function $\psi_n - k_n$ if necessary, we can assume without loss of generality that $\psi_n(0) = 0$ for all $n \in [1 \ldots N]$. Now let $s := \psi_1(1)$. By replacing ψ_n with the function ψ_n/s for all $n \in [1 \ldots N]$ if necessary, we can assume without loss of generality that $\psi_1(1) = 1$.

For every $N \in \mathbb{N}$, we can repeat the above construction. That is, for all $N \in \mathbb{N}$, we obtain a collection of functions $\psi_1^N, \ldots, \psi_N^N : \mathbb{R} \longrightarrow \mathbb{R}$ yielding an additive representation (B15) for $\succeq_N$, and furthermore such that $\psi_1^N(1) = 1$ and $\psi_n^N(0) = 0$ for all $n \in [1 \ldots N]$. ◇ Claim 6

Now I will show that these additive representations agree for different values of N.

Claim 7 *There is single infinite sequence of functions* $(\phi_n)_{n=1}^{\infty}$ *such that*

$$\psi_n^N = \phi_n, \quad \textit{for all } N \in \mathbb{N} \textit{ and all } n \in [1 \ldots N]. \tag{B17}$$

Proof Let $N, M \in \mathbb{N}$, and let $n \in [1 \ldots N]$. I will show that $\psi_n^N = \psi_n^M$. Let $a, b \in \mathbb{R}$, let $m \in [1 \ldots N]$ with $m \neq n$, and suppose there exist some $c, d \in \mathbb{R}$ such that $\psi_n^N(b) - \psi_n^N(a) = \psi_m^N(d) - \psi_m^N(c)$. (Such a c and d always exist if a and b are close enough together.) Thus, if $\mathbf{x} \in \mathcal{X}_N$ is any social outcome such that $x_n^{\uparrow} = a$ and $x_m^{\uparrow} = c$, and $b_{(n)}\mathbf{x}$ and $d_{(m)}\mathbf{x}$ are well defined, then the additive representation (B15) yields $b_{(n)}\mathbf{x} \approx d_{(m)}\mathbf{x}$. Thus, $(a \overset{n}{\rightsquigarrow} b) \cong (c \overset{m}{\rightsquigarrow} d)$.

Now, for any $M, L \in \mathbb{N}$, find $\mathbf{y} \in \mathcal{X}_M$ and $\mathbf{z} \in \mathcal{X}_L$ such that $y_n^{\uparrow} = a$ and $z_m^{\uparrow} = c$, $b_{(n)}\mathbf{y}$ and $d_{(m)}\mathbf{z}$ are well defined, and $\mathbf{y} \approx \mathbf{z}$. Then Trade-off Consistency says that $b_{(n)}\mathbf{y} \approx d_{(m)}\mathbf{z}$. Thus, $\psi_n^M(b) - \psi_n^M(a) = \psi_m^L(d) - \psi_m^L(c)$. Note that this equation holds for any $L \in \mathbb{N}$. In particular, it holds for $L = N$; thus, $\psi_n^M(b) - \psi_n^M(a) = \psi_m^N(d) - \psi_m^N(c)$. But $\psi_m^N(d) - \psi_m^N(c) = \psi_n^N(b) - \psi_n^N(a)$ by construction of c and d. Thus, we conclude that $\psi_n^M(b) - \psi_n^M(a) = \psi_n^N(b) - \psi_n^N(a)$.

This argument works for any sufficiently close $a, b \in \mathbb{R}$. Thus, for any $a \in \mathbb{R}$, there is some $\epsilon > 0$ such that $\psi_n^M(b) - \psi_n^M(a) = \psi_n^N(b) - \psi_n^N(a)$ for all $b \in (a-\epsilon, a+\epsilon)$. Since $\mathbb{R}$ can be covered with overlapping intervals like this, we conclude that $\psi_n^M(b) - \psi_n^M(a) = \psi_n^N(b) - \psi_n^N(a)$ for all $a, b \in \mathbb{R}$. Thus, there is some constant $k \in \mathbb{R}$ such that $\psi_n^M = \psi_n^N + k$. But $\psi_n^M(0) = 0 = \psi_n^N(0)$ by the construction in Claim 6. Thus, $k = 0$. Thus, $\psi_n^M = \psi_n^N$.

This argument works for all $N < M \in \mathbb{N}$ and all $n \in [1 \ldots N]$. ◇ Claim 7

For any $\mathbf{x} \in \mathcal{X}_{\propto}$, we define $\Phi(\mathbf{x}) := \sum_{n=1}^{N} \phi_n(x_n^{\uparrow})$, where $N := |\mathbf{x}|$. For any $\mathbf{x}, \mathbf{y} \in \mathcal{X}_{\propto}$ with $|\mathbf{x}| = |\mathbf{y}|$, Claims 6 and 7 together imply that

$$\big(\mathbf{x} \succeq \mathbf{y}\big) \iff \big(\Phi(\mathbf{x}) \geqslant \Phi(\mathbf{y})\big). \tag{B18}$$

It remains to show that statement (B18) also holds when $|\mathbf{x}| \neq |\mathbf{y}|$. For any $M, N \in \mathbb{N}$, let $\mathcal{I}_{M,N} := \{r \in \mathbb{R};\ \text{there exist } \mathbf{x} \in \mathcal{X}_N \text{ and } \mathbf{y} \in \mathcal{X}_M \text{ such that } \mathbf{x} \approx \mathbf{y} \text{ and } \Phi(\mathbf{x}) = r\}$.

Claim 8 *$\mathcal{I}_{M,N}$ is an nonempty interval. Thus, for any $r \in \mathbb{R}$, if $r \notin \mathcal{I}_{M,N}$, then either $r < s$ for all $s \in \mathcal{I}_{M,N}$, or $r > s$ for all $s \in \mathcal{I}_{M,N}$. In particular, for any $\mathbf{y} \in \mathcal{X}_N$,*

(a) $\big(\Phi(\mathbf{y}) < s \textit{ for all } s \in \mathcal{I}_{M,N}\big) \iff \big(\mathbf{y} \prec \mathbf{z} \textit{ for all } \mathbf{z} \in \mathcal{X}_M\big)$.

(b) $\big(\Phi(\mathbf{y}) > s \textit{ for all } s \in \mathcal{I}_{M,N}\big) \iff \big(\mathbf{y} \succ \mathbf{z} \textit{ for all } \mathbf{z} \in \mathcal{X}_M\big)$.

Proof *Nonempty* For any $N \in \mathbb{N}$, Neutral population growth yields some $\mathbf{x}_N \in \mathcal{X}_N$ such that $\mathbf{x}_N \approx \emptyset$. Let $s := \Phi(\mathbf{x}_N)$. Then $s \in \mathcal{I}_{M,N}$, because $\mathbf{x}_N \approx \mathbf{x}_M$, and $\mathbf{x}_M \in \mathcal{X}_M$.

Interval Let $r, t \in \mathcal{I}_{M,N}$, with $r < t$. We claim that $[r, t] \subseteq \mathcal{I}_{M,N}$. To see this, let $s \in (r, t)$. There exists some $\mathbf{x}, \mathbf{z} \in \mathcal{X}_N$ such that $\Phi(\mathbf{x}) = r$ and $\Phi(\mathbf{z}) = t$, and such that $\mathbf{x} \approx \mathbf{x}'$ and $\mathbf{z} \approx \mathbf{z}'$ for some $\mathbf{x}', \mathbf{z}' \in \mathcal{X}_M$. Define $\Phi_N : \mathbb{R}^{N\uparrow} \longrightarrow \mathbb{R}$ by setting $\Phi_N(\mathbf{y}) := \sum_{n=1}^{N} \phi_n(y_n)$ for all $\mathbf{y} = (y_1, \ldots, y_N) \in \mathbb{R}^{N\uparrow}$; then, Φ_N is continuous (because each of $\phi_1, \ldots, \phi_N$ is continuous). Since $\Phi_N(\mathbf{x}^\uparrow) = r$ and $\Phi_N(\mathbf{z}^\uparrow) = t$, and $\mathbb{R}^{N\uparrow}$ is connected, the intermediate value theorem yields some $\mathbf{v} \in \mathbb{R}^{N\uparrow}$ such that $\Phi_N(\mathbf{v}) = s$. Let $\mathbf{y} \in \mathcal{X}_N$ such that $\mathbf{y}^\uparrow = \mathbf{v}$; then, $\Phi(\mathbf{y}) = s$. By statement (B18), we have $\mathbf{x} \prec \mathbf{y} \prec \mathbf{z}$, because $r < s < t$.

Let $\mathcal{A} := \{\mathbf{a}^\uparrow;\ \mathbf{a} \in \mathcal{X}_M \text{ and } \mathbf{a} \succ \mathbf{y}\}$ and $\mathcal{B} := \{\mathbf{b}^\uparrow;\ \mathbf{b} \in \mathcal{X}_M \text{ and } \mathbf{b} \prec \mathbf{y}\}$. By the axiom Continuity, these are both open subsets of $\mathbb{R}^{M\uparrow}$. Clearly, they are disjoint. Furthermore, both are nonempty, because $(\mathbf{x}')^\uparrow \in \mathcal{B}$ and $(\mathbf{z}')^\uparrow \in \mathcal{A}$ (because $\mathbf{x}' \approx \mathbf{x} \prec \mathbf{y}$ and $\mathbf{z}' \approx \mathbf{z} \succ \mathbf{y}$). Thus, there must be some $(\mathbf{y}')^\uparrow \in \mathbb{R}^{M\uparrow}$ such that $\mathbf{y}' \approx \mathbf{y}$—otherwise, $\mathbb{R}^{M\uparrow} = \mathcal{A} \sqcup \mathcal{B}$, which contradicts the fact that $\mathbb{R}^{M\uparrow}$ is connected. Since $s = \Phi(\mathbf{y})$ and $\mathbf{y} \approx \mathbf{y}'$, it follows that $s \in \mathcal{I}_{M,N}$, as desired. This argument works for any $r, t \in \mathcal{I}_{M,N}$ and $s \in [r, t]$; it follows that $\mathcal{I}_{M,N}$ is an interval.

(a) "$\Longrightarrow$" (by contradiction) Let $\mathbf{y} \in \mathcal{X}_N$, and suppose $\Phi(\mathbf{y}) < s$ for all $s \in \mathcal{I}_{M,N}$, but also suppose $\mathbf{y} \succeq \mathbf{z}'$ for some $\mathbf{z}' \in \mathcal{X}_M$. Now, $\mathcal{I}_{M,N}$ is nonempty, so let $s \in \mathcal{I}_{M,N}$, and let $\mathbf{x} \in \mathcal{X}_N$ such that $\Phi(\mathbf{x}) = s$. We have $\Phi(\mathbf{y}) < s = \Phi(\mathbf{x})$, and hence, $\mathbf{y} \prec \mathbf{x}$ by statement (B18). Meanwhile, there is some $\mathbf{x}' \in \mathcal{X}_M$ such that $\mathbf{x} \approx \mathbf{x}'$, by definition of $\mathcal{I}_{M,N}$. Thus, $\mathbf{y} \prec \mathbf{x}'$. Meanwhile, $\mathbf{y} \succeq \mathbf{z}'$. By repeating the argument in the previous paragraph (using Continuity and the connectedness of $\mathbb{R}^{M\uparrow}$), we can construct some $\mathbf{y}' \in \mathcal{X}_M$ such that $\mathbf{y} \approx \mathbf{y}'$. But then $\Phi(\mathbf{y}) \in \mathcal{I}_{M,N}$, which is a contradiction. To avoid the contradiction, we must have $\mathbf{y} \prec \mathbf{z}'$.
"$\Longleftarrow$" Suppose $\mathbf{y} \prec \mathbf{z}$ for all $\mathbf{z} \in \mathcal{X}_M$. Let $s \in \mathcal{I}_{M,N}$. Then $s = \Phi(\mathbf{x})$ for some $\mathbf{x} \in \mathcal{X}_N$, with some $\mathbf{x}' \in \mathcal{X}_M$ such that $\mathbf{x} \approx \mathbf{x}'$. But then $\mathbf{y} \prec \mathbf{x}'$; hence, $\mathbf{y} \prec \mathbf{x}$; hence, $\Phi(\mathbf{y}) < \Phi(\mathbf{x}) = s$, by statement (B18), as desired.

The proof of (b) is very similar to the proof of (a). ◇ Claim 8

For any $r \in \mathcal{I}_{M,N}$, find $\mathbf{x} \in \mathcal{X}_N$ such that $\Phi(\mathbf{x}) = r$. Then find $\mathbf{y} \in \mathcal{X}_M$ with $\mathbf{x} \approx \mathbf{y}$, and define $V_{N,M}(r) := \Phi(\mathbf{y})$. Then $V_{N,M}(r) \in \mathcal{I}_{N,M}$.

Claim 9 *$V_{N,M}(r)$ is well defined independent of the particular choice of* $\mathbf{x}$ *and* $\mathbf{y}$.

Proof Let $\mathbf{x}' \in \mathcal{X}_N$ and $\mathbf{y}' \in \mathcal{X}_M$, and suppose that $\Phi(\mathbf{x}') = r$ and $\mathbf{x}' \approx \mathbf{y}'$. Then $\mathbf{y}' \approx \mathbf{x}' \approx \mathbf{x} \approx \mathbf{y}$ (where the middle indifference is by (B18), because $\Phi(\mathbf{x}') = r = \Phi(\mathbf{x})$) hence $\mathbf{y}' \approx \mathbf{y}$ (by transitivity), and hence $\Phi(\mathbf{y}') = \Phi(\mathbf{y})$ [by (B18)]. ◇ Claim 9

This yields a function $V_{M,N} : \mathcal{I}_{M,N} \longrightarrow \mathcal{I}_{N,M}$. It is easily verified that $V_{M,N}$ is an increasing bijection from $\mathcal{I}_{M,N}$ to $\mathcal{I}_{N,M}$, and $V_{M,N}^{-1} = V_{N,M}$, as a function from $\mathcal{I}_{N,M}$ back to $\mathcal{I}_{M,N}$.

Claim 10 *For any* $\mathbf{x} \in \mathcal{X}_N$ *and* $\mathbf{y} \in \mathcal{X}_M$*, if* $\Phi(\mathbf{x}) \in \mathcal{I}_{M,N}$*, then*

$$\big(\mathbf{x} \succeq \mathbf{y}\big) \iff \big(V_{M,N}[\Phi(\mathbf{x})] \geqslant \Phi(\mathbf{y})\big).$$

Proof Let $r := \Phi(\mathbf{x})$, and let $r' := V_{M,N}(r)$. Then there is some $\mathbf{x}' \in \mathcal{X}_M$ such that $\mathbf{x} \approx \mathbf{x}'$ and $\Phi(\mathbf{x}') = r'$. Let $s := \Phi(\mathbf{y})$. If $s \leqslant r'$, then representation (B18) yields $\mathbf{y} \preceq \mathbf{x}'$. Meanwhile, $\mathbf{x}' \approx \mathbf{x}$; thus, $\mathbf{y} \preceq \mathbf{x}$, by transitivity. If $s \geqslant r'$, then representation (B18) yields $\mathbf{y} \succeq \mathbf{x}'$. Meanwhile, $\mathbf{x}' \approx \mathbf{x}$; thus, $\mathbf{y} \succeq \mathbf{x}$, by transitivity. ⋄ Claim 10

Claim 11 *For any* $n < m \in \mathbb{N}$ *and* $a < c \in \mathbb{R}$*, there exists* $\epsilon > 0$ *and a continuous, increasing function* $\psi : (a - \epsilon, a + \epsilon) \longrightarrow \mathbb{R}$ *with* $\psi(a) = c$*, such that for all* $b \in (a - \epsilon, a + \epsilon)$*, if* $d := \psi(b)$*, then* $(a \overset{n}{\rightsquigarrow} b) \cong (c \overset{m}{\rightsquigarrow} d)$*.*

Proof Let $\mathbf{x} \in \mathcal{X}_\propto$ such that $x^\uparrow_{n-1} < x^\uparrow_n = a < x^\uparrow_{n+1}$ and $x^\uparrow_{m-1} < x^\uparrow_m = c < x^\uparrow_{m+1}$. Let $N := |\mathbf{x}|$. Since ϕ_m is continuous and strictly increasing, its image $\mathcal{R}_m := \phi_m(\mathbb{R})$ is an open interval in $\mathbb{R}$, and $\phi_m : \mathbb{R} \longrightarrow \mathcal{R}_m$ is a homeomorphism. Likewise, if $\mathcal{R}_n := \phi_n(\mathbb{R})$, then $\mathcal{R}_n$ is an open interval and $\phi_n : \mathbb{R} \longrightarrow \mathcal{R}_n$ is a homeomorphism. Let $\mathcal{R}'_m := \{r - \phi_m(c) + \phi_n(a); r \in \mathcal{R}_m\}$; then, $\phi_n(a) \in \mathcal{R}'_m$ (because $\phi_m(c) \in \mathcal{R}_m$), and thus, $\mathcal{R}'_m \cap \mathcal{R}_n$ is itself a nonempty open interval containing $\phi_n(a)$. Let $\mathcal{Q}_n := \phi_n^{-1}(\mathcal{R}'_m \cap \mathcal{R}_n)$; then, $\mathcal{Q}_n$ is an open interval containing a. Now define $\psi : \mathcal{Q}_n \longrightarrow \mathbb{R}$ by setting

$$\psi(q) := \phi_m^{-1}\left(\phi_n(q) - \phi_n(a) + \phi_m(c)\right), \quad \text{for all } q \in \mathcal{Q}_n.$$

Then $\psi(a) = c$. If $\mathcal{Q}_m := \psi(\mathcal{Q}_n)$, then $\mathcal{Q}_m$ is an open interval containing c, and ψ is a continuous, increasing bijection from $\mathcal{Q}_n$ to $\mathcal{Q}_m$. Let $\mathcal{Q}'_n := \mathcal{Q}_n \cap (x^\uparrow_{n-1}, x^\uparrow_{n+1}) \cap \psi^{-1}(x^\uparrow_{m-1}, x^\uparrow_{m+1})$, and let $\mathcal{Q}'_m := \psi(\mathcal{Q}'_n)$, then $\mathcal{Q}'_n$ and $\mathcal{Q}'_m$ are open intervals around a and c, respectively, and $\psi : \mathcal{Q}'_n \longrightarrow \mathcal{Q}'_m$ is a continuous, increasing function.

For any $b \in \mathcal{Q}'_n$, the element $b_{(n)}\mathbf{x}$ is well defined because $x^\uparrow_{n-1} < b < x^\uparrow_{n+1}$. If $d := \psi(b)$, then $d_{(m)}\mathbf{x}$ is well defined because $x^\uparrow_{m-1} < d < x^\uparrow_{m+1}$ because $d \in \mathcal{Q}'_m$. Finally, $b_{(n)}\mathbf{x} \approx d_{(m)}\mathbf{x}$ by statement (B18), because $\Phi(b_{(n)}\mathbf{x}) = \Phi(d_{(m)}\mathbf{x})$, because $\phi_m(d) - \phi_m(c) = \phi_n(b) - \phi_n(a)$ by the definition of ψ. Since $x^\uparrow_n = a$ and $x^\uparrow_m = c$, we conclude that $(a \overset{n}{\rightsquigarrow} b) \cong (c \overset{m}{\rightsquigarrow} d)$.

Now find $\epsilon > 0$ small enough that $(a - \epsilon, a + \epsilon) \subseteq \mathcal{Q}'_n$. Then for any $b \in (a - \epsilon, a + \epsilon)$, if $d = \psi(b)$, then $(a \overset{n}{\rightsquigarrow} b) \cong (c \overset{m}{\rightsquigarrow} d)$, by the previous paragraph. ⋄ Claim 11

Claim 12 *For any* $M, N \in \mathbb{N}$*, there exists a constant* $Q_{M,N} \in \mathbb{R}$ *such that* $V_{M,N}(r) = r + Q_{M,N}$ *for all* $r \in \mathcal{I}_{M,N}$*.*

Proof Let $r \in \mathcal{I}_{M,N}$. Find $\mathbf{x} \in \mathcal{X}_N$ with $\Phi(\mathbf{x}) = r$, and find $\mathbf{y} \in \mathcal{X}_M$ such that $\mathbf{x} \approx \mathbf{y}$; then, $V_{M,N}[\Phi(\mathbf{x})] = \Phi(\mathbf{y})$, by the definition of $V_{M,N}$ and Claim 9. Find $n, m \in [1 \ldots N]$ such that $x_{n-1} < x_n < x_{n+1}$ and $y_{m-1} < y_m < y_{m+1}$. Let $a := x_n$ and $c := y_m$. Let $\psi : (a - \epsilon, a + \epsilon) \longrightarrow \mathbb{R}$ be as described in Claim 11; then, $\psi(a) = c$. Define

$$\epsilon_0 := \min\left\{\epsilon,\ x_{n+1} - a,\ a - x_{n-1},\ \psi^{-1}(y_{m+1}) - a,\ a - \psi^{-1}(y_{m-1})\right\}.$$

Then $\epsilon_0 > 0$. Let $b \in (a - \epsilon_0, a + \epsilon_0)$, and let $d := \psi(b)$. If $\delta := \phi_n(b) - \phi_n(a)$, then also $\phi_m(d) - \phi_m(c) = \delta$, because $(a \overset{n}{\rightsquigarrow} b) \cong (c \overset{m}{\rightsquigarrow} d)$ by the definition of ψ in Claim 11. If $\mathbf{x}' := b_{(n)}\mathbf{x}$ (which is well defined because $b \in (x_{n-1}, x_{n+1})$), then $\Phi(\mathbf{x}') = \Phi(\mathbf{x}) + \delta$. Likewise, if $\mathbf{y}' := d_{(m)}\mathbf{y}$ [which is well defined because $d = \psi(b)$ and $b \in (\psi^{-1}(y_{m-1}),\ \psi^{-1}(y_{m+1}))$], then $\Phi(\mathbf{y}') = \Phi(\mathbf{y}) + \delta$.

As $\mathbf{x} \approx \mathbf{y}$, $\mathbf{x}' \approx \mathbf{y}'$, by Trade-off Consistency. Thus, $V_{M,N}[\Phi(\mathbf{x}')] = \Phi(\mathbf{y}')$, by Claim 10. In other words, $V_{M,N}[\Phi(\mathbf{x}) + \delta] = \Phi(\mathbf{y}) + \delta = V_{M,N}[\Phi(\mathbf{x})] + \delta$.

This equality holds for any sufficiently small δ—in particular, it holds for all δ in the set $\{\phi_n(b) - \phi_n(a);\, b \in (a - \epsilon_0, a + \epsilon_0)\}$, which is an open interval around zero. Thus, if $r \in \mathcal{I}_{M,N}$ and $s \in \mathcal{I}_{N,M}$ are any values such that $V_{M,N}(r) = s$, then we also have $V_{M,N}(r + \delta) = s + \delta$ for all sufficiently small δ. This shows that $V_{M,N}$ is an affine function with slope 1 in a neighbourhood of each point in $\mathcal{I}_{M,N}$. But $\mathcal{I}_{M,N}$ is an interval by Claim 8; it follows that $V_{M,N}$ is an affine function with slope 1 everywhere on $\mathcal{I}_{M,N}$. ◇ Claim 12

Based on Claim 12, we can extend $V_{M,N}$ to an affine function $V_{M,N} : \mathbb{R} \longrightarrow \mathbb{R}$, by defining $V_{M,N}(r) = r + Q_{M,N}$ for all $r \in \mathbb{R}$.

Claim 13 *For any* $\mathbf{x} \in \mathcal{X}_N$ *and* $\mathbf{z} \in \mathcal{X}_M$, $\big(\mathbf{x} \preceq \mathbf{z}\big) \iff \big(V_{M,N}[\Phi(\mathbf{x})] \leqslant \Phi(\mathbf{z})\big)$.

Proof Let $r := \Phi(\mathbf{x})$ and let $t := \Phi(\mathbf{z})$. If $r \in \mathcal{I}_{M,N}$, then the stated equivalence follows from Claim 10. Likewise, if $t \in \mathcal{I}_{N,M}$, then it follows from Claim 10 and the observation that $V_{N,M}^{-1} = V_{M,N}$ and both are increasing, so that $V_{M,N}[\Phi(\mathbf{x})] \leqslant \Phi(\mathbf{z})$ if and only if $\Phi(\mathbf{x}) \leqslant V_{N,M}[\Phi(\mathbf{z})]$.

So, suppose that $r \notin \mathcal{I}_{M,N}$ and $t \notin \mathcal{I}_{N,M}$. It follows that $\mathbf{x} \not\approx \mathbf{z}$ (because otherwise we would have both $r \in \mathcal{I}_{M,N}$ and $t \in \mathcal{I}_{N,M}$). Thus, either $\mathbf{x} \prec \mathbf{z}$ or $\mathbf{x} \succ \mathbf{z}$.

Claim 13A (a) *If* $\mathbf{x} \prec \mathbf{z}$, *then* $V_{M,N}[\Phi(\mathbf{x})] < \Phi(\mathbf{z})$.
(b) *If* $\mathbf{x} \succ \mathbf{z}$, *then* $V_{M,N}[\Phi(\mathbf{x})] > \Phi(\mathbf{z})$.

Proof (a) Suppose $\mathbf{x} \prec \mathbf{z}$. Claim 8 says $\mathcal{I}_{M,N}$ is an interval. So, since $r \notin \mathcal{I}_{M,N}$, we must have either $r < s$ for all $s \in \mathcal{I}_{M,N}$, or $r > s$ for all $s \in \mathcal{I}_{M,N}$. If $r > s$ for all $s \in \mathcal{I}_{M,N}$, then Claim 8(b) says that $\mathbf{x} \succ \mathbf{y}$ for all $\mathbf{y} \in \mathcal{X}_M$, which contradicts the hypothesis that $\mathbf{x} \prec \mathbf{z}$. So, we must have $r < s$ for all $s \in \mathcal{I}_{M,N}$. By a similar logic [using Claim 8(a)], we must have $t > s'$ for all $s' \in \mathcal{I}_{N,M}$.

Now, let $s \in \mathcal{I}_{M,N}$ and find some $\mathbf{y} \in \mathcal{X}_N$ such that $\Phi(\mathbf{y}) = s$, and some $\mathbf{y}' \in \mathcal{X}_M$ such that $\mathbf{y} \approx \mathbf{y}'$. Thus, if $s' := \Phi(\mathbf{y}')$, then $s' = V_{M,N}(s)$. Furthermore, $s' \in \mathcal{I}_{N,M}$. By the previous paragraph, we have $r < s$ and $s' < t$. Thus, $V_{N,M}(r) < V_{N,M}(s) = s' < t$. In other words, $V_{M,N}[\Phi(\mathbf{x})] < \Phi(\mathbf{z})$.

The proof of (b) is similar. ▽ Claim 13A

Claim 13B $V_{M,N}[\Phi(\mathbf{x})] \neq \Phi(\mathbf{z})$.

Proof (by contradiction) Suppose $V_{M,N}[\Phi(\mathbf{x})] = \Phi(\mathbf{z})$. By taking the contrapositive parts (a) and (b) of Claim 13A, we cannot have either $\mathbf{x} \prec \mathbf{z}$ or $\mathbf{x} \succ \mathbf{z}$. So we must have $\mathbf{x} \approx \mathbf{z}$, because $\succeq$ is a complete relation. But we have already deduced that $\mathbf{x} \not\approx \mathbf{z}$, so this is a contradiction. ▽ Claim 13B

It follows from Claim 13B that either $V_{M,N}[\Phi(\mathbf{x})] < \Phi(\mathbf{z})$ or $V_{M,N}[\Phi(\mathbf{x})] > \Phi(\mathbf{z})$. If $V_{M,N}[\Phi(\mathbf{x})] < \Phi(\mathbf{z})$, then the contrapositive of Claim 13A(b) says that $\mathbf{x} \preceq \mathbf{z}$, and hence, $\mathbf{x} \prec \mathbf{z}$ (because $\mathbf{x} \not\approx \mathbf{z}$). If $V_{M,N}[\Phi(\mathbf{x})] > \Phi(\mathbf{z})$, then the contrapositive of Claim 13A(a) says that $\mathbf{x} \succeq \mathbf{z}$, and hence, $\mathbf{x} \succ \mathbf{z}$ (because $\mathbf{x} \not\approx \mathbf{z}$). At this point, we have shown that $\mathbf{x} \prec \mathbf{z}$ if and only if $V_{M,N}[\Phi(\mathbf{x})] < \Phi(\mathbf{z})$. Likewise, $\mathbf{x} \succ \mathbf{z}$, if and only if $V_{M,N}[\Phi(\mathbf{x})] > \Phi(\mathbf{z})$. Since we also know that $\mathbf{x} \not\approx \mathbf{z}$ and $V_{M,N}[\Phi(\mathbf{x})] \neq \Phi(\mathbf{z})$ (by Claim 13B), this suffices to prove the claimed equivalence. ◇ Claim 13

For all $N, M \in \mathbb{N}$, let $Q_{N,M}$ be as in Claim 12.

Claim 14 *For all $M, N \in \mathbb{N}$, we have $Q_{M,N} = -Q_{N,M}$, and for all $L \in \mathbb{N}$, we have $Q_{L,M} + Q_{M,N} = Q_{L,N}$.*

Proof As already noted, $V_{M,N}^{-1} = V_{N,M}$, as a function from $\mathcal{I}_{N,M}$ back to $\mathcal{I}_{M,N}$; thus, Claim 12 yields $Q_{M,N} = -Q_{N,M}$.

Now consider the set $\mathcal{I}_{M,N} \cap V_{N,M}(\mathcal{I}_{L,M})$. I claim this intersection is nonempty. To see this, for all $\ell \in \{L, M, N\}$, let $\mathbf{x}_\ell \in \mathcal{X}_\ell$ be such that $\mathbf{x}_\ell \approx \emptyset$; such elements exist by Neutral population growth. If $r := \Phi(\mathbf{x}_N)$, then $r \in \mathcal{I}_{M,N}$ (because $\mathbf{x}_N \approx \mathbf{x}_M$). Likewise, if $s := \Phi(\mathbf{x}_M)$, then $s \in \mathcal{I}_{L,M}$ (because $\mathbf{x}_M \approx \mathbf{x}_L$). Finally, $V_{N,M}(s) = r$, because $\mathbf{x}_M \approx \mathbf{x}_N$. Thus, $r \in \mathcal{I}_{M,N} \cap V_{N,M}(\mathcal{I}_{L,M})$; thus, $\mathcal{I}_{M,N} \cap V_{N,M}(\mathcal{I}_{L,M}) \neq \{\}$.

It is easily verified that $\mathcal{I}_{M,N} \cap V_{N,M}(\mathcal{I}_{L,M}) \subseteq \mathcal{I}_{L,N}$, and $V_{L,M} \circ V_{M,N}(r) = V_{L,N}(r)$, for all $r \in \mathcal{I}_{M,N} \cap V_{N,M}(\mathcal{I}_{L,M})$. Thus, Claim 12 yields $Q_{L,M} + Q_{M,N} = Q_{L,N}$. ◇ Claim 14

For all $N \in \mathbb{N}$, let $q_N := Q_{N,N-1}$. (In particular, $q_1 = Q_{1,0} = V_{1,0}(0) = V_{1,0}[\Phi(\emptyset)] = \phi_1(x_1)$, where $x_1 \in \mathbb{R}$ is the unique value such that if $\mathbf{x} \in \mathcal{X}_1$ is the one-person outcome with lifetime utility x_1, then $\mathbf{x} \approx \emptyset$; such an x_1 exists by Neutral population growth, and it is unique by Pareto.) For any $N < M$, Claim 14 implies that $Q_{M,N} = q_{N+1} + \cdots + q_M$. For all $n \in \mathbb{N}$, define $\phi'_n := \phi_n - q_n$. For any $\mathbf{x} \in \mathcal{X}_\infty$, if $N := |\mathbf{x}|$, then define

$$\Phi'(\mathbf{x}) := \sum_{n=1}^{N} \phi'_n(x_n^{\uparrow}) = \sum_{n=1}^{N} \phi_n(x_n^{\uparrow}) - \sum_{n=1}^{N} q_n = \Phi(\mathbf{x}) - Q_{N,0}. \tag{B19}$$

Thus, for all $M \in \mathbb{N}$ and $\mathbf{y} \in \mathcal{X}_M$,

$$\big(\Phi'(\mathbf{x}) \geqslant \Phi'(\mathbf{y})\big) \underset{\text{(a)}}{\iff} \big(\Phi(\mathbf{x}) - Q_{N,0} \geqslant \Phi(\mathbf{y}) - Q_{M,0}\big)$$
$$\iff \big(\Phi(\mathbf{x}) + Q_{M,0} - Q_{N,0} \geqslant \Phi(\mathbf{y})\big) \underset{\text{(b)}}{\iff} \big(V_{M,N}[\Phi(\mathbf{x})] \geqslant \Phi(\mathbf{y})\big) \underset{\text{(c)}}{\iff} (\mathbf{x} \succeq \mathbf{y}),$$

as desired. Here, (a) is by Eq. (B19). Next, (b) is because $Q_{M,0} - Q_{N,0} = Q_{M,N}$ by Claim 14, so that $\Phi(\mathbf{x}) + Q_{M,0} - Q_{N,0} = \Phi(\mathbf{x}) + Q_{M,N} = V_{M,N}[\Phi(\mathbf{x})]$. Finally, (c) is by Claim 13. □

Remarks In the proof of Theorem 2, Neutral population growth is only needed in Claims 8 and 14, where it is used to show that certain sets are not empty.

If we had assumed that $\succeq$ satisfied an axiom of Separability similar to the one used in Sect. 2.2, then Claims 1–5 would be unnecessary, and the proof of Claim 6 could be made much simpler: we could just invoke Corollary 3.6 of Wakker (1993) to immediately obtain an additive representation of $\succeq_N$ on all of $\mathbb{R}^{N}{\uparrow}$. (This argument was used in an earlier version of this paper.)

Proof of Proposition 3.1 "(a) $\Longrightarrow$ (b)" (by contradiction) Let $I := \inf(\phi_1(\mathbb{R}))$. If statement (b) is false, then either $I > -S(\boldsymbol{\phi})$, or $I = -S(\boldsymbol{\phi})$ and supremum in formula (3F) *is* obtained.

Case 1 Suppose $I > -S(\boldsymbol{\phi})$. Then $-I < S(\boldsymbol{\phi})$. Thus, there exist $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_N \in \mathbb{R}$ such that $\sum_{n=1}^{N} \delta\phi_n(x_n) > -I$. Find $\mathbf{x} \in \mathcal{X}_\propto$ such that $\mathbf{x}^\uparrow = (x_1, \ldots, x_N)$. Suppose $r < x_1$, and let $\mathbf{y} := \mathbf{x} \uplus r$. Then $\mathbf{y}^\uparrow = (r, x_1, \ldots, x_N)$. Thus,

$$W(\mathbf{y}) = \phi_1(r) + \sum_{n=1}^{N} \phi_{n+1}(x_n), \text{ while } W(\mathbf{x}) \;=\; \sum_{n=1}^{N} \phi_n(x_n),$$

$$\text{so that } W(\mathbf{y}) - W(\mathbf{x}) = \phi_1(r) + \sum_{n=1}^{N} \delta\phi_n(x_n) > \phi_1(r) - I \underset{(*)}{\geqslant} 0,$$

where $(*)$ is by definition of I. Thus, $W(\mathbf{y}) > W(\mathbf{x})$, so $\mathbf{x} \uplus r \succ \mathbf{x}$. This holds for all $r < x_1$.

On the other hand, if $s \geqslant x_1$, then $s > r$ for any $r < x_1$, and thus $\mathbf{x} \uplus s \succ \mathbf{x} \uplus r$ by Pareto, while $\mathbf{x} \uplus r \succ \mathbf{x}$ by the previous paragraph. Thus, $\mathbf{x} \uplus s \succ \mathbf{x}$ by transitivity. It follows that $\mathbf{x} \uplus s \succ \mathbf{x}$ for *all* $s \in \mathbb{R}$. This contradicts the axiom Critical levels.

Case 2 Suppose $I = -S(\boldsymbol{\phi})$ and supremum in formula (3F) *is* obtained. Then $-I = S(\boldsymbol{\phi})$, and there exists some $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_N \in \mathbb{R}$ such that $\sum_{n=1}^{N} \delta\phi_n(x_n) = -I$. Again, let $\mathbf{x} \in \mathcal{X}_\propto$ be such that $\mathbf{x}^\uparrow = (x_1, \ldots, x_N)$, let $r < x_1$, and let $\mathbf{y} := \mathbf{x} \uplus r$. Then by a similar computation to *Case 1*, we get

$$W(\mathbf{y}) - W(\mathbf{x}) \;=\; \phi_1(r) + \sum_{n=1}^{N} \delta\phi_n(x_n) = \phi_1(r) - I > 0.$$

(Here, the last step is because $\phi_1(r) > I$ because the infimum I is *never* obtained, since ϕ_1 is strictly increasing.) Thus, once again, $W(\mathbf{y}) > W(\mathbf{x})$; hence, $\mathbf{x} \uplus r \succ \mathbf{x}$. This argument holds for all $r < x_1$. The rest of the argument is identical to *Case 1*; again, we obtain a contradiction of Critical levels.

"(b) $\Longleftarrow$ (a)" Suppose $\succeq$ has an ARA representation satisfying the condition the theorem. To show that $\succeq$ satisfies Critical levels, let $\mathbf{x} \in \mathcal{X}_\propto$. For any $r \in \mathbb{R}$, define $\psi(r) = W(\mathbf{x} \uplus r)$. It is easily verified that $\psi : \mathbb{R} \longrightarrow \mathbb{R}$ is a continuous function. To verify Critical levels, we must find some $c \in \mathbb{R}$ such that $\psi(c) = W(\mathbf{x})$.

Claim 1 *There exists $d \in \mathbb{R}$ such that $\psi(d) > W(\mathbf{x})$.*

Proof Let $N := |\mathbf{x}|$, and let $\mathbf{x}^\uparrow = (x_1^\uparrow, \ldots, x_N^\uparrow)$. By hypothesis, we have $\phi_{N+1}(c_{N+1}) = 0$. Thus, $\phi_{N+1}(d) > 0$ for any $d > c_{N+1}$. Suppose $d > \max\{x_N^\uparrow, c_{N+1}\}$, and let $\mathbf{d} := \mathbf{x} \uplus d$. Then $\mathbf{d}^\uparrow = (x_1^\uparrow, \ldots, x_N^\uparrow, d)$. Thus, $\psi(d) = W(\mathbf{d}) = W(\mathbf{x}) + \phi_{N+1}(d) > W(\mathbf{x})$, because $\phi_{N+1}(d) > 0$. $\diamond$ Claim 1

Claim 2 *There exists $b \in \mathbb{R}$ with $\phi_1(b) < -\sum_{n=1}^{N} \delta\phi_n(x_n^{\uparrow})$.*

Proof Let $A := \sum_{n=1}^{N} \delta\phi_n(x_n^{\uparrow})$. Then $S(\boldsymbol{\phi}) \geqslant A$, so $-S(\boldsymbol{\phi}) \leqslant -A$. By hypothesis, $\inf(\phi_1(\mathbb{R})) \leqslant -S(\boldsymbol{\phi})$, and if $\inf(\phi_1(\mathbb{R})) = -S(\boldsymbol{\phi})$, then the supremum (3F) is not obtained. If $\inf(\phi_1(\mathbb{R})) < -S(\boldsymbol{\phi})$, then there is some $b \in \mathbb{R}$ such that $\phi_1(b) < -S(\boldsymbol{\phi})$, and hence, $\phi_1(b) < -A$ as desired. On the other hand, if $\inf(\phi_1(\mathbb{R})) = -S(\boldsymbol{\phi})$, then the supremum (3F) is not obtained, so $S(\phi) > A$. Thus, $-S(\phi) < -A$, and hence $\inf(\phi_1(\mathbb{R})) < -A$, so there is some $b \in \mathbb{R}$ such that $\phi_1(b) < -A$, as desired.

◇ Claim 2

Claim 3 *There exists $b \in \mathbb{R}$ with $\psi(b) < W(\mathbf{x})$.*

Proof Let $b_0 \in \mathbb{R}$ be as in Claim 2. Note that any $b < b_0$ also satisfies the inequality in Claim 2. By making b small enough, we can assume that $b < x_1^{\uparrow}$. Thus, if $\mathbf{b} = \mathbf{x} \uplus b$, then $\mathbf{b}^{\uparrow} = (b, x_1^{\uparrow}, \ldots, x_N^{\uparrow})$. Thus,

$$W(\mathbf{b}) = \phi_1(b) + \sum_{n=1}^{N} \phi_{n+1}(x_n^{\uparrow}), \text{ while } W(\mathbf{x}) = \sum_{n=1}^{N} \phi_n(x_n^{\uparrow}),$$

$$\text{so that } W(\mathbf{b}) - W(\mathbf{x}) = \phi_1(b) + \sum_{n=1}^{N} \delta\phi_n(x_n^{\uparrow}) < 0.$$

Thus, $\psi(b) = W(\mathbf{b}) < W(\mathbf{x})$. ◇ Claim 3

From Claims 1 and 3, we have $b, d \in \mathbb{R}$ such that $\psi(b) < W(\mathbf{x}) < \psi(d)$. By the intermediate value theorem, there exists some $c \in (b, d)$ such that $\psi(c) = W(\mathbf{x})$. Thus, $W(\mathbf{x} \uplus c) = W(\mathbf{x})$, which means $\mathbf{x} \uplus c \approx \mathbf{x}$, as desired. □

Proof of Proposition 3.2 The proof of (a) is similar to the proof of Proposition 2.2(a). The proof of parts (b) and (c) is similar to the proof of Proposition 2.4. □

Proof of Proposition 3.4 (a) Instead of Minimal Aggregation, we will actually show that any ARA SWO satisfies the following, slightly stronger axiom:

> Minimal Aggregation$^+$ For any $N \in \mathbb{N}$, any $\alpha > 0$ and all $\mathbf{x} \in \mathcal{X}_N$, there exists some $\beta > 0$ such that, for all $\mathbf{y} \in \mathcal{X}_\propto$ with $\mathcal{I}(\mathbf{y}) = \mathcal{I}(\mathbf{x})$, if there is some $i \in \mathcal{I}(\mathbf{x})$ such that $x_i \geqslant y_i \geqslant x_i - \beta$, while $y_j \geqslant x_j + \alpha$ for all other $j \in \mathcal{I}(\mathbf{x})$, then $\mathbf{y} \succeq \mathbf{x}$.

If we represent $\succeq$ with an ordering $\succeq_*$ on $\mathbb{R}^{\propto\uparrow}$, as in formula (3A), then this axiom is equivalent to:

> Minimal Aggregation* For any $N \in \mathbb{N}$, any $\alpha > 0$ and all $\mathbf{x} \in \mathbb{R}^{N\uparrow}$, there exists some $\beta > 0$ such that, for all $\mathbf{y} \in \mathbb{R}^N$, if there is some $n \in [1 \ldots N]$ such that $x_n \geqslant y_n \geqslant x_n - \beta$, while $y_n \geqslant x_n + \alpha$ for all other $n \in [1 \ldots N]$, then $\mathbf{y}^{\uparrow} \succeq_* \mathbf{x}$.

(Here, $\mathbf{y}^{\uparrow}$ is the vector obtained by reordering the components of $\mathbf{y}$ in increasing order.)

To prove this, let $N \in \mathbb{N}$, and $\mathbf{x} \in \mathbb{R}^{N\uparrow}$. Let $N_1 < N_2 < \cdots N_J \leqslant N$ be the unique values such that

$$x_1 = \cdots = x_{N_1-1} < x_{N_1} = \cdots = x_{N_2-1} < x_{N_2} = \cdots < \cdots$$
$$= x_{N_J-1} < x_{N_J} = \cdots = x_N.$$

(For example, if $x_1 < x_2 < \cdots < x_N$, then we have $J = N - 1$, with $N_1 = 2$, $N_2 = 3, \cdots, N_J = N$). Let $\overline{\alpha} := \min\{x_{N_j} - x_{N_j-1};\ j \in [1 \ldots J]\}$; then, $\overline{\alpha} > 0$. Fix $\alpha > 0$. If the axiom holds for α, then it holds for any $\alpha' > \alpha$ (by Pareto). Thus, we can assume that $\alpha < \overline{\alpha}$ without loss of generality. Let $\overline{\beta} := \overline{\alpha} - \alpha$; then, $\overline{\beta} > 0$. For any $n \in [1 \ldots N]$, let

$$M_n := \sum_{\substack{m=1\\ m\neq n}}^{N} \left(\phi_m(x_m + \alpha) - \phi(x_m)\right).$$

Then $M_n > 0$ because $\phi_1, \ldots, \phi_N$ are all strictly increasing. Let $M := \min\{M_1, \ldots, M_N\}$. For all $n \in [1 \ldots N]$, there exists some $\beta_n > 0$ small enough that $\phi_n(x_n) - \phi_n(x_n - \beta_n) < M$ (because ϕ_n is continuous). Let $\beta := \min\{\alpha, \overline{\beta}, \beta_1, \ldots \beta_N\}$; then, $\alpha > \beta > 0$.

Let $\mathbf{y} \in \mathbb{R}^N$, let $n \in [1 \ldots N]$, and suppose $x_n > y_n \geqslant x_n - \beta$, while $y_m \geqslant x_m + \alpha$ for all $m \neq n$. There are now two cases to consider:

Case 1 Suppose $y_m < x_m + \overline{\alpha}$ for all $m \neq n$. Then for all $j \in [1 \ldots J]$, all $n < N_j$, and all $m \geqslant N_j$, we have $y_n < y_m$. Thus, by permuting the coordinates internally within each of the blocks $[1 \ldots N_1)$, $[N_1 \ldots N_2)$, $[N_2 \ldots N_3)$, $\cdots$, $[N_J \ldots N)$ (which does not change $\mathbf{x}$, by definition of $N_1, \ldots, N_J$), we can assume without loss of generality that $y_1 \leqslant y_2 \leqslant \cdots \leqslant y_N$, so that $\mathbf{y}^{\uparrow} = \mathbf{y}$. Furthermore,

$$\begin{aligned}\phi_n(x_n) - \phi_n(y_n) &\leqslant \phi_n(x_n) - \phi_n(x_n - \beta) \leqslant \phi_n(x_n) - \phi_n(x_n - \beta_n)\\ &< M \leqslant M_n = \sum_{\substack{m=1\\ m\neq n}}^{N} \left(\phi_m(x_m + \alpha) - \phi(x_m)\right)\\ &\leqslant \sum_{\substack{m=1\\ m\neq n}}^{N} \left(\phi_m(y_m) - \phi(x_m)\right).\end{aligned}$$

Rearranging this inequality, we get $W(\mathbf{y}) \geqslant W(\mathbf{x})$, and thus, $\mathbf{y} \succeq_* \mathbf{x}$, as desired.

Case 2 Suppose $y_m \geqslant x_m + \overline{\alpha}$ for some $m \neq n$. Define $\mathbf{z}$ by setting $z_n := y_n$ and $z_m := \min\{y_m, x_m + \alpha\}$ for all $m \neq n$. Then $y_m \geqslant z_m$ for all $m \in [1 \ldots N]$. Thus, $y_m^{\uparrow} \geqslant z_m^{\uparrow}$ for all $m \in [1 \ldots N]$, so $\mathbf{y}^{\uparrow} \succeq_* \mathbf{z}^{\uparrow}$ by the Pareto axiom. Meanwhile, $\mathbf{z}$ satisfies *Case 1* (because $\alpha < \overline{\alpha}$), so $\mathbf{z}^{\uparrow} = \mathbf{z} \succeq_* \mathbf{x}$. By transitivity, we get $\mathbf{y}^{\uparrow} \succeq_* \mathbf{x}$, as desired.

(b) Let $r, q \in \mathbb{R}$ with $r > q$. Let $\alpha := (r - q)/2$. By hypothesis, there is some $A > 0$ such that $\sum_{n=N+1}^{\infty} \phi_n'(q + \alpha) = A \cdot \phi_m'(q + \alpha)$, for all $m, N \in \mathbb{N}$ with $m \leqslant N$. Since the functions $\{\phi_n'\}_{n=1}^{\infty}$ are all nonincreasing (by concavity), this means that

$$\sum_{n=N+1}^{\infty} \phi_n'(x_*) \leqslant A \cdot \phi_m'(q+\alpha), \text{ for all } x_* \geqslant q+\alpha \text{ and } m, N \in \mathbb{N} \text{ with } m \leqslant N. \tag{B20}$$

Let $\beta > 0$. For any $n \in \mathbb{N}$, since ϕ_n is concave, we have

$$\phi_n(x_* + \beta) - \phi_n(x_*) \leqslant \beta \cdot \phi_n'(x_*), \quad \text{for all } n \geqslant N+1. \tag{B21}$$

In particular, if $\beta \leqslant \alpha/A$, then we can substitute inequality (B21) into inequality (B20) to obtain:

$$\sum_{n=N+1}^{\infty} \big(\phi_n(x_* + \beta) - \phi(x_*)\big) \leqslant \alpha \cdot \phi_m'(q+\alpha), \text{ for all } m, N \in \mathbb{N} \text{ with } m \leqslant N. \tag{B22}$$

Meanwhile, if $y \leqslant q$, then

$$\phi_m(y+\alpha) - \phi_m(y) \geqslant \alpha \cdot \phi_m'(y+\alpha) \geqslant \alpha \cdot \phi_m'(q+\alpha), \tag{B23}$$

because ϕ_m' is concave. Combining inequalities (B22) and (B23) yields

$$\sum_{n=N+1}^{\infty} \big(\phi_n(x_* + \beta) - \phi(x_*)\big) \leqslant \phi_m(y+\alpha) - \phi_m(y), \tag{B24}$$

for any $x_* \geqslant q+\alpha$ and $y \leqslant q$, and any $m, N \in \mathbb{N}$ with $m \leqslant N$.

Let $\beta < \min\{\alpha, \alpha/A\}$; then, $r - \beta > q + \alpha$ (because $\alpha = (r-q)/2$). Let $\mathbf{x}, \mathbf{y} \in \mathcal{X}_\propto$ with $\mathcal{I}(\mathbf{x}) = \mathcal{I}(\mathbf{y}) = \mathcal{J}$, for some finite subset $\mathcal{J} \subset \mathcal{I}$. Suppose $i \in \underline{\mathcal{I}}(\mathbf{y})$ and $y_i \leqslant q$, and $x_i \geqslant y_i + \alpha$, while for all other $j \in \mathcal{J}$, suppose that either $x_j = y_j$ or $j \in \overline{\mathcal{I}}(\mathbf{x}) \cap \overline{\mathcal{I}}(\mathbf{y})$, $y_j \geqslant r$ and $x_j \geqslant y_j - \beta$. In particular, this means there exists some $y_* \geqslant r$ and $x_* \geqslant y_* - \beta > q + \alpha$ such that $y_j = y_*$ and $x_j = x_*$ for all $j \in \overline{\mathcal{I}}(\mathbf{x}) \cap \overline{\mathcal{I}}(\mathbf{y})$. Let $N := \big|\mathcal{J} \setminus \big(\overline{\mathcal{I}}(\mathbf{x}) \cap \overline{\mathcal{I}}(\mathbf{y})\big)\big|$; thus, in both $\mathbf{x}$ and $\mathbf{y}$, all individuals in $\overline{\mathcal{I}}(\mathbf{x}) \cap \overline{\mathcal{I}}(\mathbf{y})$ have rank at least $N+1$ when all members of $\mathcal{J}$ are ordered from lowest to highest utility. Furthermore, since all these individuals have identical utility in $\mathbf{x}$, and have identical utility in $\mathbf{y}$, we can assume without loss of generality that they have the *same* rank in $\mathbf{x}$ and $\mathbf{y}$. For all $j \in \overline{\mathcal{I}}(\mathbf{x}) \cap \overline{\mathcal{I}}(\mathbf{y})$, let $n(j)$ be the rank that we assign to j in $\mathbf{x}$ and $\mathbf{y}$; thus, $n(j) \geqslant N+1$. Thus,

$$\begin{aligned}
\sum_{j \in \overline{\mathcal{I}}(\mathbf{x}) \cap \overline{\mathcal{I}}(\mathbf{y})} \big(\phi_{n(j)}(y_j) - \phi_{n(j)}(x_j)\big) &= \sum_{j \in \overline{\mathcal{I}}(\mathbf{x}) \cap \overline{\mathcal{I}}(\mathbf{y})} \big(\phi_{n(j)}(y_*) - \phi_{n(j)}(x_*)\big) \\
\underset{(*)}{\leqslant} \sum_{j \in \overline{\mathcal{I}}(\mathbf{x}) \cap \overline{\mathcal{I}}(\mathbf{y})} \big(\phi_{n(j)}(x_* + \beta) - \phi_{n(j)}(x_*)\big) &\underset{(\dagger)}{\leqslant} \sum_{n=N+1}^{\infty} \big(\phi_n(x_* + \beta) - \phi_n(x_*)\big) \\
\underset{(\diamond)}{\leqslant} \phi_m(y_i + \alpha) - \phi_m(y_i) &\underset{(\ddagger)}{\leqslant} \phi_m(x_i) - \phi_m(y_i), \text{ for all } m \in [1 \ldots N].
\end{aligned} \tag{B25}$$

Here, $(*)$ is because $x_* \geqslant y_* - \beta$, $(\dagger)$ is because $n(j) \geqslant N+1$ for all $j \in \overline{\mathcal{I}}(\mathbf{x}) \cap \overline{\mathcal{I}}(\mathbf{y})$, $(\diamond)$ is by inequality (B24), because $y_i \leqslant q$, and $(\ddagger)$ is because $x_i \geqslant y_i + \alpha$.

Let $m := |\underline{\mathcal{I}}(\mathbf{y})|$; then, $m \leqslant N$, by the definition of N. There are now two cases.

Case 1 Suppose that $x_i \leqslant x_j$, for all $j \in \mathcal{J} \setminus \overline{\mathcal{I}}(\mathbf{y})$. Thus, in outcome $\mathbf{x}$, individual i has rank $m \leqslant N$ when all members of $\mathcal{J}$ are ordered from lowest to highest utility. In outcome $\mathbf{y}$, individual i would have rank *at most* m, but since the bottom m individuals have identical utilities, we can assume without loss of generality that i also has rank m in $\mathbf{y}$. By hypothesis, we have $x_j = y_j$ for all $j \in \mathcal{J} \setminus \left(\{i\} \cup \left[\overline{\mathcal{I}}(\mathbf{x}) \cap \overline{\mathcal{I}}(\mathbf{y})\right]\right)$. Thus,

$$W(\mathbf{x}) - W(\mathbf{y}) = \left(\phi_m(x_i) - \phi_m(y_i)\right) - \sum_{j \in \overline{\mathcal{I}}(\mathbf{x}) \cap \overline{\mathcal{I}}(\mathbf{y})} \left(\phi_{n(j)}(y_j) - \phi_{n(j)}(x_j)\right) \underset{(*)}{\geqslant} 0,$$

where $(*)$ is by inequality (B25). Thus, $\mathbf{x} \succeq \mathbf{y}$, as claimed

Case 2 Suppose that $x_i > x_j$, for some $j \in \mathcal{J} \setminus \overline{\mathcal{I}}(\mathbf{y})$. In this case, define $\mathbf{z}$ by setting $z_i := \min\{x_j;\ j \in \mathcal{J} \setminus \underline{\mathcal{I}}(\mathbf{y})\}$, while $z_j := x_j$ for all other $j \in \mathcal{I}$. Then $\mathbf{x}$ Pareto dominates $\mathbf{z}$, so $\mathbf{x}^{\uparrow}$ Pareto dominates $\mathbf{z}^{\uparrow}$, so $\mathbf{x} \succeq \mathbf{z}$ by the Pareto axiom. Meanwhile, $\mathbf{z}$ satisfies *Case 1*, so $\mathbf{z} \succeq \mathbf{y}$. Thus, by transitivity, $\mathbf{x} \succeq \mathbf{y}$, as desired. □

Proof of Corollary 3.5 The strategy is very similar to the proof of Proposition 2.1. The uniqueness statement follows from the uniqueness statement in Theorem 2. □

References

Adler, M.D.: Future generations: a prioritarian view. George Wash. Law Rev. **77**, 1478–1520 (2008)

Adler, M.D.: Claims across outcomes and population ethics. In: Arrhenius and Bykvist (2019)

Arrhenius, G.: An impossibility theorem for welfarist axiologies. Econ. Philos. **16**(2), 247–266 (2000)

Arrhenius, G.: Population Ethics. Oxford University Press, Oxford (2018)

Arrhenius, G., Bykvist, K. (eds.): The Oxford Handbook of Population Ethics. Oxford University Press, Oxford (2019). **(forthcoming)**

Arrhenius, G., Rabinowicz, W.: Better to be than not to be? In: Joas, H. (ed.) Benefit of Broad Horizons: Intellectual and Institutional Preconditions for a Global Social Science, pp. 399–421. Leiden, Brill (2010)

Arrhenius, G., Rabinowicz, W.: The value of existence. In: Hirose, I., Olson, J. (eds.) The Oxford Handbook of Value Theory, pp. 424–43. Oxford University Press, Oxford (2015)

Arrhenius, G., Ryberg, J., Tännsjö, T.: The repugnant conclusion. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University (2019)

Asheim, G.B., Zuber, S.: Escaping the repugnant conclusion: rank-discounted utilitarianism with variable population. Theor. Econ. **9**(3), 629–650 (2014)

Asheim, G.B., Zuber, S.: Evaluating intergenerational risks. J. Math. Econ. **65**, 104–117 (2016)

Asheim, G., Zuber, S.: Rank-discounting as a resolution to a dilemma in population ethics. (2017) **(preprint)**

Balasubramanian, A.: On weighted utilitarianism and an application. Soc. Choice Welf. **44**(4), 745–763 (2015)

Bernoulli, D.: Exposition of a new theory on the measurement of risk. Econometrica **22**(1), 23–36. (Translated by Louise Sommer. Original title: Specimen Theoriae Novae de Mensura Sortis, 1738 (1954))

Blackorby, C., Donaldson, D.: Ratio-scale and translation-scale full interpersonal comparability without domain restrictions: admissible social-evaluation functions. Int. Econ. Rev. **23**(2), 249–268 (1982)

Blackorby, C., Bossert, W., Donaldson, D.: Population ethics and the existence of value functions. J. Public Econ. **82**(2), 301–308 (2001)

Blackorby, C., Bossert, W., Donaldson, D.J.: Population Issues in Social Choice Theory, Welfare Economics, and Ethics. Cambridge University Press, Cambridge (2005)

Bleichrodt, H., Rohde, K.I., Wakker, P.P.: Combining additive representations on subsets into an overall representation. J. Math. Psychol. **52**(5), 304–310 (2008)

Bossert, W.: An axiomatization of the single-series Ginis. J. Econ. Theory **50**(1), 82–92 (1990)

Chateauneuf, A., Wakker, P.: From local to global additive representation. J. Math. Econ. **22**, 523–545 (1993)

Cowen, T.: Resolving the repugnant conclusion. In: Ryberg and Tännsjö, pp. 81–98 (2004)

Debreu, G.: Representation of a preference ordering by a numerical function. In: Thrall, R., Coombs, C., Davis, R. (eds.) Decision Processes, pp. 159–165. Wiley, New York (1954)

Debreu, G.: Topological methods in cardinal utility theory. In: Arrow, K.J., Karlin, S., Suppes, P. (eds.) Mathematical Methods in the Social Sciences 1959, pp. 16–26. Stanford University Press, Stanford (1960)

Donaldson, D., Weymark, J.A.: A single-parameter generalization of the Gini indices of inequality. J. Econ. Theory **22**(1), 67–86 (1980)

Ebert, U.: Measurement of inequality: an attempt at unification and generalization. Soc. Choice Welf. **5**, 147–169 (1988)

Fleurbaey, M., Tungodden, B.: The tyranny of non-aggregation versus the tyranny of aggregation in social choices: a real dilemma. Econ. Theory **44**(3), 399–414 (2010)

Fleurbaey, M., Voorhoeve, A.: On the social and personal value of existence. In: Hirose, I., Reisner, A. (eds.) Weighing and Reasoning: Themes from the Philosophy of John Broome, pp. 95–109. Oxford University Press, Oxford (2015)

Greaves, H.: Population axiology. Philos. Compass **12**(11), e12442 (2017)

Hare, C.: Voices from another world: must we respect the interests of people who do not, and will never, exist? Ethics **117**(3), 498–523 (2007)

Holtug, N.: On the value of coming into existence. J. Ethics **5**(4), 361–384 (2001)

McMahan, J.: Problems of population theory. Ethics **92**(1), 96–127 (1981)

McMahan, J.: Asymmetries in the morality of causing people to exist. In: Roberts, M.A., Wasserman, D.T. (eds.) Harming Future Persons, pp. 49–68. Springer, Dordrecht (2009)

Mongin, P., Pivato, M.: Ranking multidimensional alternatives and uncertain prospects. J. Econ. Theory **157**, 146–171 (2015)

Mongin, P., Pivato, M.: Social evaluation under risk and uncertainty. In: Adler, M.D., Fleurbaey, M. (Eds.) Handbook of Well-being and Public Policy, Ch. 24, pp. 711–745. Oxford University Press, Oxford (2016)

Mongin, P., Pivato, M.: Social preference under twofold uncertainty (2018) **(preprint)**

Nozick, R.: Anarchy, State, and Utopia. Basic Books, New York (1974)

Parfit, D.: Reasons and Persons. Oxford University Press, Oxford (1984)

Roberts, M.A.: Can it ever be better never to have existed at all? Person-based consequentialism and a new repugnant conclusion. J. Appl. Philos. **20**(2), 159–185 (2003)

Roberts, M.A.: The asymmetry: a solution. Theoria **77**(4), 333–367 (2011)

Roberts, M.A.: The nonidentity problem. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University (2019)

Roemer, J.E.: Eclectic distributional ethics. Polit. Philos. Econ. **3**(3), 267–281 (2004)

Ryberg, J., Tännsjö, T. (eds.): The Repugnant Conclusion: Essays on Population Ethics. Springer, New York (2004)

Sider, T.R.: Might theory X be a theory of diminishing marginal value? Analysis **51**(4), 265–271 (1991)

Thomas, T.: Separability. In: Arrhenius and Bykvist (2019) **(to appear)**

Wakker, P.: The algebraic versus the topological approach to additive representations. J. Math. Psych. **32**(4), 421–435 (1988)

Wakker, P.: Additive representations on rank-ordered sets. II. The topological approach. J. Math. Econ. **22**, 1–26 (1993)

Weymark, J.A.: Generalized Gini inequality indices. Math. Soc. Sci. **1**(4), 409–430 (1981)

Yaari, M.E.: A controversial proposal concerning inequality measurement. J. Econ. Theory **44**(2), 381–397 (1988)

Zank, H.: Social welfare functions with a reference income. Soc. Choice Welf. **28**(4), 609–636 (2007)