Population Ethics: A Position-Based Axiomatic Approach*

Alyssa Rusonik† ⓡ and Itzhak Gilboa‡

October 27, 2025

Abstract

A social planner considers the far future, asking which population profile should be preferred, where none of the people involved has been born or conceived yet. A population's profile is given by the number of individuals who are in each possible position. Thus, symmetry among individuals who are in the same position is presupposed by the model. The model allows populations to be of different sizes, and assumes that they can be compared by the social planner. Three simple conditions characterize the relations that can be represented in a utilitarian way, that is, by assigning a number to each position so that profiles are ranked according to the sum of utilities across individuals.

1 Introduction

What are our preferences for the profile of humanity 200 years hence? Do we prefer a population of 10 billion people living in comfort, or 20 billion who suffer hunger? Or, given a population size, do we prefer that they work hard and

*We thank Loren Fryxell, Margaret Meyer, and Peter Wakker for comments and references. Gilboa gratefully acknowledges Investissements d'Avenir ANR -11- IDEX-0003 / Labex ECODEC No. ANR - 11-LABX-0047, the AXA Chair for Decision Sciences at HEC. Rusonik gratefully acknowledges a state grant managed by the Agence Nationale de la Recherche under the Investissements d'Avenir programme (ANR-18-EURE-0005 / EUR DATA EFM)

†HEC, Paris. alyssa.rusonik@hec.edu

‡HEC, Paris. gilboa@hec.fr

be healthy, or need not work but cope with diseases? Such questions, which are within the realm of population ethics, have received growing attention since the works of Parfit (1976, 1982) and they seem to become more pertinent every decade, when humanity faces challenges such as climate change, as well as the opportunities and risks presented by technological changes.

By allowing population sets to differ, population ethics engages with conceptual questions that go beyond the standard problems arising in social choice theory which takes the set of agents as given. The so-called "Repugnant Conclusion," for instance, introduced by Parfit (1984), observes that pure utilitarianism – and indeed many aggregative welfare approaches – would inevitably prefer a sufficiently-large, marginally-subsistent society to a small, flourishing one.[1]

The theory of social choice provides conceptual frameworks within which this and other moral puzzles can be formulated and addressed. Among the most influential is Critical Level Utilitarianism (Blackorby, Bossert, and Donaldson, 1995; 1997): by positing the existence of a critical threshold, below which individuals' well-being does not contribute to aggregate welfare, Critical Level Utilitarianism allows for a comparison of welfare of population groups of different sizes. A strictly-positive critical threshold avoids the "Repugnant Conclusion", while a critical threshold of zero does not. In Critical Level Utilitarianism, the "critical level" is not unique and its interpretation can depend on the precise application.[2]

The present note offers a new framework to address the problem. It differs from the standard approaches in abstracting away from individual identities, with a degree of anonymity built into the model. We assume as primitive a set of "positions" people may occupy. It is implicitly assumed that a "position" contains all the relevant information for the social choice problem. A

[1] Also known as the "Mere Addition Paradox", the paradoxical nature of the problem is contested: some philosophers have argued that the finding is merely counterintuitive and have disputed the "repugnancy" of the Conclusion (Huemer, 2008; Tännsjö , 2002; Gustaffson, 2022), while others have considered this as revelatory of a fundamental flaw to aggregation of welfare in ethics and utilitarianism (Temkin, 2012).

[2] See Broome (2004) for a discussion of the merits and drawbacks of aggregative welfare approaches in general, and non-unique critical thresholds in particular.

population's profile is a "counter vector" I, where $I(p)$ indicates how many individuals are at position p. The overall number of individuals in a profile (ranging over all positions) is finite, but we impose no a priori bounds – neither on the number of positions which may be occupied nor on the number of individuals at each position. We assume that a social planner can rank such profiles, and show that a few simple axioms on the ranking are equivalent to the existence of a representation thereof by a utilitarian aggregation function.

Our axiomatization derives a unique zero-threshold, uniform across individuals, which arises naturally from the underlying preferences. One of the innovations of our approach is to start with positions: this allows us to avoid axioms that explicitly guarantee the uniqueness and uniformity of the threshold. Our model is best interpreted as describing a social planner who gauges the utility of individuals based on criteria she deems salient, without soliciting the individuals' subjective assessments of their own positions. The interpretation of the zero-threshold is, then, similar to one accommodated by some versions of Critical Level Utilitarianism: the zero-level is assigned exactly to those positions for which (the social planner believes) individuals are indifferent between existence and non-existence.

Importantly, we do not necessarily view the result as an argument in favor of utilitarianism. Rather, we consider it as an exercise that can shed some new light on the concept. We tend to believe that many would find utilitarian aggregation reasonable in some settings but not in others, and the axiomatic derivation suggested here may help one draw the line between the two.

The axioms we use are the following. First, preferences over profiles are assumed to be a weak order. Second, they should satisfy a "Union" axiom: if the planner prefers profile I to J, she should prefer $(I + K)$ to $(J + K)$. Recalling that the profiles keep count of how many individuals are at each position, the profile $(I + K)$ corresponds to the union of two disjoint populations, one with profile I, and the other – with K. Thus, the Union axiom could be illustrated as follows: suppose that the social planner can choose between one policy, leading to a population profile I, and another, which will result in a population profile J. It turns out that there is another location, where the population profile will be K, irrespective of the current policy choice. The Union axiom states that the preference between I and J should not change if each is added the same population, whose profile is K.

Our third and last axiom is Archimedean: it states that, should profile K be considered strictly better than L, then, no matter which populations I, J we start out with, taking their union with sufficiently many copies of K and L, respectively, would result in a preference for the I-with-K's population over the J-with-L's one. As usual, it is an axiom of commensurability: it states that all preferences can be compared, and that none is infinitely more weighty than another. While such an axiom is often considered "technical", in our setup we consider it key. For example, we may agree that a population with a single individual who is alive (J) is preferred to a population with a single individual who is dead (I). At the same time, we may also agree that a population with a single individual enjoying a monthly income of \$1,100 is happier than a population with a single individual whose monthly income is only \$1,000. The Archimedean axiom implies that we should prefer a population with a dead individual, should this be compensated by sufficiently many people being richer. If, however, extreme outcomes are excluded, the axiom may appear less objectionable. In fact, we view it as a key axiom in judging the scope of utilitarian reasoning. We discuss this point below.

Taken together, the three axioms are equivalent to the existence of a number $u(p)$ for each position p, such that any two profiles are ranked according to the summation of $u(p)$ over all individuals in the profile. These numbers are unique up to a multiplicative constant. While the set of positions may be infinite, every profile considered has only finitely many individuals, so that the summation of their utilities is well-defined. It is possible, for example, that a position is defined by a level of wealth, and that the set of positions is a continuum. Yet, a given population will only have finitely many individuals, and therefore only finitely many positions p such that $I(p) > 0$. The theorem states that there exists a number $u(p)$ for each p, such that population profile I is ranked according to $\sum I(p) u(p)$ and thus $u(p)$ can be viewed as the utility function from wealth. There is, however, no assumption about the utility

u as a function of p; the mathematical derivation uses the space of counter vectors (the I, J...) and not the utility values or even the positions.

Observe that an assumption of anonymity is built into our model: individuals' identities are ignored, and only the number of individuals in each position is taken into account. However, the notion of a "position" is rather abstract, and it is up to the modeler to decide what information is encoded in positions. For example, if a position is defined by monthly income and family size, the model can be viewed as anonymous. If, however, a proper name is part of the information embedded in a "position", the model hardly satisfies anonymity in any intuitive sense.

The utility function we derive is unique only up to the choice of a unit of measurement, and one does not have the freedom to shift it by an additive constant, that is, to select the location of zero. As mentioned above, this is in line with the notion of a unique critical threshold in population ethics. We may consider a society I and ask how it compares with the same society, when we add one individual at position p. The answer would reveal the sign of $u(p)$. Intuitively, the zero value on the utility scale can be viewed as the threshold below which (the social planner thinks that) it is better not to exist than to exist. More generally, the utility function will not be invariant to additive shifts even if it is always positive, because it provides answers to questions such as "Will you consider the society better off if it has n individuals in position p or m individuals in position q?"

Section 2 proceeds with a survey of the literature. Section 3 describes the model and result formally. Section 4 deals with special cases and extensions of the basic model, exploiting the freedom allowed by the notion of a "position". It is followed by a discussion in Section 5, while the proof of the representation theorem is given in an appendix.

2 Literature Review

A philosophical schism divides the study of population ethics into two main fields of thought: person-affecting views and impersonal approaches. Person-affecting theories (see, e.g., Navreson, 1973, for a foundational work) maintain that an outcome can only be better or worse if it is better or worse for some particular individual, leading to the "non-identity problem" and to difficulties in evaluating the moral significance of creating new lives (Parfit, 1984). Impersonal approaches, conversely, focus on aggregate welfare measures independent of the identities of the individuals. They often lead to counterintuitive or potentially-problematic implications like the "Repugnant Conclusion". Seminal philosophical works on the subject include those by Carlson (1998), Arrhenius (2000), Roberts (2011), Boonin (2014), and Frick (2020). The issue has also been discussed in social choice theory by Blackorby and Donaldson (1984), among others.[3]

At first sight, this paper's axiomatization appears firmly rooted in the impersonal approach, due to its commitment to agents' anonymity. Indeed, in our model only the number of individuals in each position counts. The individuals in our model have no names or identity numbers. There is no reference to their subjective utility functions. However, as mentioned above, the notion of a "position" is open to a variety of interpretations. While the entire model is viewed in the eyes of a social planner, this social planner may allow individual idiosyncrasies to be part of the definition of positions.

As for aggregative approaches to ethics and the aforementioned "Repugnant Conclusion" in particular, a number of theoretical responses have been proposed, including variants of average utilitarianism[4] (Harsanyi, 1955; Sikora, 1975); rank-dependent approaches (Zuber and Asheim, 2012); prioritarianism (Adler, 2012; Parfit, 1997); and various threshold-based modifications, includ-

[3]This philosophical dichotomy is strongly evocative of a debate at the intersection of logic and linguistics. Narveson (1973) juxtaposes Bertrand Russell's claim (see, e.g., Quine (1962, p. 221)) that sentences whose subjects contain references to the non-existent are inherently false (if "they" do not exist, no true assertion can be made about "their" rights, so any statement is a priori false) with Strawson (1950), which introduces the notion of presupposition: a view that sentences by construction assume the referent's existence (if the "they" in question do exist, the statement makes logical sense and has a truth value, and it is in this context that the statement should be read; and if "they" were, in fact, non-existent, the statement would be assigned no truth value at all).

[4]Which, nonetheless, are exposed to critiques analogous to the "Repugnant Conclusion;" see also Parfit (1984) and Arrhenius (2000).

ing leximin principles (Hammond, 1976), utilitarianism dampened by population (Hurka, 1983), and maximin welfare orderings (Bossert, 1990). Our paper can be viewed as adding to this literature by deriving an analog to the critical threshold organically and in guaranteeing its uniqueness, both in level and across individuals.

3 Model and Result

Let there be a set of *positions* P. The set of *profiles* is

$$\mathcal{I} = \left\{ I : P \to \mathbb{Z}_+ \;\middle|\; \sum_{p \in P} I(p) < \infty \right\}$$

For $I \in \mathcal{I}$, the *support* of I is $supp(I) \equiv \{ p \in P \mid I(p) > 0 \}$. Thus, the set of positions P may be infinite, but each profile I has a finite support. Algebraic operations are performed on $\mathcal{I}$ pointwise so that $nI + mJ$ is well-defined for $n, m \geq 0$ and $I, J \in \mathcal{I}$. $0 \in \mathcal{I}$ has its natural meaning. Similarly, for $I, J \in \mathcal{I}$ the inequality $I \geq J$ is read pointwise.

Let there be given a binary relation $\succsim \subset \mathcal{I} \times \mathcal{I}$. The symmetric and asymmetric parts of $\succsim$ will be denoted, as usual, by $\sim$ and $\succ$, respectively. Define:

A1 Weak Order: $\succsim$ is complete and transitive.

A2 Union: For all $I, J, K \in \mathcal{I}$, $I \succsim J$ iff $I + K \succsim J + K$.

A3 Archimedeanity: For all $I, J, K, L \in \mathcal{I}$, if $K \succ L$, then there exists $n \geq 0$ such that $I + nK \succ J + nL$.

Theorem 1 *$\succsim$ satisfies A1-A3 iff there exists a utility function*

$$u : P \to \mathbb{R}$$

such that, for all $I, J \in \mathcal{I}$

$$\begin{array}{ccc} I & \succsim & J \\ & \textit{iff} & \\ \sum_{p \in P} I(p)\, u(p) & \geq & \sum_{p \in P} J(p)\, u(p) \end{array} \quad (1)$$

Moreover, in this case u is unique up to multiplication by a positive number.

From a mathematical viewpoint, this result is similar to de Finetti's (1931, 1937) derivation of maximization of expected value relative to a subjective probability (see the version described in Gilboa, 2009). In that model, P is a set of states of the world, and a real-valued function from the states to the real line describes a bet. Similar axioms are used to derive a separating hyperplane between the bets that are at least as desirable as zero and those that are not, and an additional monotonicity axiom guarantees that (apart from the trivial case of indifference), the gradient of the hyperplane can be interpreted as a probability over the states. The major conceptual difference between the results is that in de Finetti's model the utility function is assumed known (represented by the values of each function), and the derived weights are probabilities – whereas in the present model the given values are numbers of occurrences of positions, and the utility is derived as the weights assigned to these positions. The difference in interpretation also results in a mathematical difference: in the present model the values of the functions compared are only non-negative integers, rather than any real numbers. Consequently, the proof of this result is more involved than it would be in the real-valued case.

Our paper derives a "zero" on the utility scale from the social planner's preferences. Intuitively, this is the utility assigned to a position that is neutral in the planner's eyes: adding individuals in such a position does not affect the overall evaluation of the society's well-being. This is reminiscent of other axiomatic models in which entities that are "larger" in an appropriate sense can be equivalent to "smaller" ones. For example, Gilboa and Schmeidler (2001) consider sequences of "facts" that affect a person's well-being. They allow the comparisons of sequences of different lengths, where equivalence of a sequence to a prefix thereof indicates that the suffix is "neutral". Fryxell (2024) considers a rather different model of experienced utility, involving functions that are defined on different subsets of the real line (capturing time).[5] He imposes ax-

[5]Luce and Krantz (1971) also deal with comparisons of functions defined on different domains, where their interpretation is within the theory of decision under uncertainty. They model conditional, rather than aggregate, expected utility.

ioms that guarantee time-additive aggregation with a derived utility function, which also involves a derivation of a "neutral" utility level. Finally, he also applies the model to more than a single individual, and derives utilitarianism.

4 Extensions and Special Cases

4.1 Time

Thinking of future generations, one invariably thinks about time. How long will these individuals live, and how much time will they enjoy various standards of living? Should we prefer our offspring to have long and barely-satisfactory lives, or short, pleasurable ones? Are these questions equivalent to similar ones about the size of the population at a given time, or should one treat aggregation over time differently than aggregation over individuals?

Our model suggests several natural ways in which the question of time could be addressed. First, one may interpret a position p as specifying living conditions in a timeless manner. Thus, an individual who lives k periods will be counted k times, adding to $I(p)$ the number of periods she will spend at position p. This formulation will ignore the notion of personal identity that is defined by continuity across periods. Alternatively, one may define time as part of the notion of a "position", so that an individual who lives k periods adds 1 to k different positions. In other words, we can think of a position p as a pair $p = (\hat{p}, t)$ where $\hat{p}$ describes living standards, and t – time. It would be natural then to interpret the number $u(p)$ deriving from the representation result as a product $u((\hat{p}, t)) = \hat{u}(\hat{p})\,\delta(t)$ where $\hat{u}$ is a time-independent assessment of living standards $\hat{p}$, and $\delta(t)$ is a discount factor.

At the other extreme, one may define a position to be informative enough to state an entire lifespan. Thus, a person who lives from time 50 to 150 will be counted once in a position that describes a standard of living for each of these 101 periods. A priori, this position may differ from the position describing exactly the same life course, starting at 60 and ending at 160.

In between these extremes, the social planner can choose other varieties of "positions". For example, the more informative notion of a position could be replaced by a timeless one, where a single position describes a lifespan (and its length), but not its starting time. Conversely, one may start with the first version, in which each position describes a single time period, but enrich the notion of "position" to reflect a person's memories and expectations for the future. This would allow the social planner to take into account some notions of identity, and to take into account cross-period justice considerations. For example, if we feel that an individual who survived a war deserves to be treated better than another who has had a happy past, we can introduce the distinction between them by elaborating the notion of a position, allowing the social planner to judge "living a peaceful life after the horrors of war" differently than "living a peaceful life after a happy past".

4.2 Uncertainty

Social policies determined at present affect the future, but typically not in a deterministic way. The social planner cannot precisely know what would result from her current choices, whether these are about climate change, nuclear energy, AI, or practically anything else. Our model is, however, sufficiently flexible to deal with uncertainty as well: all one needs to do is to redefine positions so that they describe a state of the world. As in the case of time periods, we may think of a position p as a pair $p = (\hat{p}, \omega)$, where $\hat{p}$ describes living standards, and ω – a state of the world. An individual who will be alive at state ω will be counted exactly once among $\{I((\hat{p}, \omega))\}_{\hat{p}}$. Correspondingly, the number $u(p)$ derived from the theorem can be thought of as $u((\hat{p}, \omega)) = \hat{u}(\hat{p})\,\pi(\omega)$ where π is interpreted as the probability of state ω.

4.3 Alternative Interpretations

The model presented here can also be used to analyze social choice with a given population at a given time. For example, assume that a position p only reflects income and that there are N individuals in a given economy. For the axiomatization to make sense, we need to assume that the social planner can compare populations of different sizes, that is, counter vectors I that do not necessarily add up to the given N. However, once the coefficients $u(p)$ have been derived under the assumption of a wider domain, they can be used to choose a maximizer of $\sum_{p\in P} I(p)\, u(p)$ under the constraint that $\sum_{p\in P} I(p) = N$. Thus, we obtain preferences over income distributions of the given population. In other words, the preferences over different population sizes are used to scale the utility values of different income levels, but they need not be used to determine the population size. A question such as "Do you think that n individuals with income x make a happier society than m individuals with income y?" can be useful to evaluate the ratio $u(x)/u(y)$, without relating to a choice problem. More concretely, a social planner may think that 100 individuals with income \$1,000 make a happier society than do 101 individuals with income \$0. This does not mean that the social planner is willing to kill one person out of 101 to make the remaining ones richer. It only says that, for that planner, $100u(100) > 101u(0)$. That is, questions about different population sizes are used to calibrate a utility function, and this function is then used for *constrained* optimization. For a given population at present, the size of the population is such a constraint.

5 Discussion

5.1 Discussion of the Axioms

Completeness: The completeness axiom is, as always, rather problematic. From a normative viewpoint, one may well pose the question, can we indeed compare all profiles? Shouldn't we perhaps be silent on some questions? Indeed, much of microeconomics focuses on Pareto domination, a notoriously (and proudly) partial order, leaving many questions unanswered, in particular those related to interpersonal comparisons of utilities.

The justification of completeness in our setup is basically identical to its justification in other models of social choice: decisions have to be taken. We may leave them outside of the model, but they won't go away, and the choices will eventually be made by politicians. Hence, the standard argument goes, it is better to explicitly model them than not. It is useful to bear in mind that a partial relation one feels strongly about may be extended in multiple ways to generate a complete relation one is not necessarily sure is the "right" choice. Yet, the argument goes, it is theoretically more satisfactory to allow for multiplicity of completions rather than to ignore them. We find that the same argument applies in our model as in other social choice models. Decisions that society takes today, whether about climate change, AI regulation, or any other problem, may imply that future societies may have this or that profile, often not necessarily with the same total number of individuals. Hence, we adopt completeness of preferences between profiles, bearing in mind that one may wish to consider partial orders, but also characterize all their completions.

While our model generates a complete ordering for the aforementioned reasons, there are papers which opt for quasi-orderings instead, including, for instance, and Parfit (1982), which introduces indeterminacy via ranges of utility specified for added individuals.

Transitivity: Transitivity is one of the most basic axioms of choice models, and we do not find that we have anything to say about it that is specific to the current model. The main reason to highlight it is that some of the literature on the "Repugnant Conclusion" suggested to drop this axiom (see Temkin, 1987, Persson, 2004, and Rachels, 2004). We tend to believe that giving up on transitivity is a high price to pay, both theoretically and practically, and that it is worthwhile to explore the possibilities that respect this axiom.

Union: The Union axiom is clearly a main assumption that drives much of the mathematical derivation of the representation. It is very similar to axioms of separability across populations, and to Savage's P2 (Savage, 1954). It also bears resemblance to vNM (1944, 1947) Independence axiom, basically stating that a subpopulation that is common to two profiles can be ignored. The axiom is surely not innocuous, and it may be violated by other-regarding preferences such as those driven by envy or empathy. Yet, when taking a normative viewpoint, one may wish to rule out other-regarding preferences and choose to satisfy the axiom as a moral principle. Be that as it may, the pros and cons of the axiom seem to be very similar to those of related axioms in other models. In a sense, dealing with yet-unborn individuals may make it easier to choose to ignore other-regarding preferences.

It should be mentioned, however, that the plausibility of the axiom depends on the exact definition of a "position" in the model. For example, one might argue that having income x while others have $2x$ is a different position to be at than having income x while all the others also have the same x. In other words, one may allow the notion of a position to include social comparisons, and emotions such as empathy or envy. In this case, some profiles may not seem feasible. Specifically, it will take some imagination to consider a society in which there are only two individuals, each at a position "enjoying income x while the other enjoys $2x$".[6]

The Union axiom thus arguably has more descriptive credibility when positions are granular enough to express the individual's role relative to that of others. Coarser position descriptions may, conversely, have normative appeal: a benevolent social planner could prefer to evaluate each individual's welfare independently, or in more objective or "real" terms, without regard to empathy or envy.

Archimedeanity: Finally, the Archimedean axiom is key, as mentioned in the introduction. It might be possible to derive versions of our main theorem that do not use this axiom, resorting to representations using non-standard analysis or lexicographically ordered utilities. Such representations may deal with the incommensurability of, say, matters of life and death with mere convenience. They are, however, beyond the scope of this paper.

6 Appendix: Proof of the Theorem

It is straightforward that the representation (1) implies that $\succsim$ satisfies A1-A3. We thus turn to prove sufficiency of the axioms, from which the uniqueness

[6]With some effort, one can suggest to consider these mental states, which might not correspond to reality because the individuals are misinformed. But this takes effort indeed.

of the representation will also follow. Naturally, the idea is to extend the relation, given for natural-number vectors, to real-valued vectors and apply a separation theorem. More specifically, we wish to successively extend $\succsim$ to vectors of integer and rational numbers, and then apply a separation theorem to the convex hull of the rational-valued vectors. To this end we define the following sets of vectors:

$$\begin{aligned}\mathcal{I}_{\mathbb{Z}} &= \{\, I : P \to \mathbb{Z} \mid \#supp\,(I) < \infty \}\\ \mathcal{I}_{\mathbb{Q}} &= \{\, I : P \to \mathbb{Q} \mid \#supp\,(I) < \infty \}\end{aligned}$$

For two sets $A \subset B$ and two binary relations on them, $\succsim \subset A \times A$ and $\succsim' \subset B \times B$, we say that $\succsim'$ is an *extension* of $\succsim$ if it extends both the weak and strict preference, that is, if, $\succsim \subset \succsim'$ and $\succ \subset \succ'$. We will also need to refer to the versions of axioms A1-A3 where the vectors are in $\mathcal{I}_{\mathbb{Z}}$ and in $\mathcal{I}_{\mathbb{Q}}$ (rather than only in $\mathcal{I}$). To simplify notation, we will use the same terms "A1", "A2", "A3", rather than redefine them for these domains.

Assume, then, that $\succsim \subset \mathcal{I} \times \mathcal{I}$ satisfies A1-A3. We first extend A2 to apply also for vectors K that are integer-valued but not necessarily nonnegative.

Lemma 1 *There exists a unique $\succsim_{\mathbb{Z}}$ on $\mathcal{I}_{\mathbb{Z}}$ satisfying A1-A3 and extending $\succsim$ (from $\mathcal{I}$ to $\mathcal{I}_{\mathbb{Z}}$).*

Proof: Let there be given $I, J \in \mathcal{I}_{\mathbb{Z}}$. Define $I \succsim_{\mathbb{Z}} J$ if there exists $K \in \mathcal{I}$ such that $I + K \succsim J + K$ (here and in the sequel, the claim that $\succsim$ holds between two vectors is taken to mean that they are both in $\mathcal{I}$ and that the relation holds between them). Clearly, if $I, J \in \mathcal{I}$ satisfy $I \succsim J$, then $I \succsim_{\mathbb{Z}} J$ (because we may use $K = 0$). Moreover, if we also have $I \succ J$ then we can conclude that $I \succ_{\mathbb{Z}} J$: otherwise, if $J \succsim_{\mathbb{Z}} I$, there would be a $K \in \mathcal{I}$ such that $J + K \succsim I + K$ and $J \succsim I$ would follow. Thus, $\succsim_{\mathbb{Z}}$ extends $\succsim$.

We turn to prove that the extended relation $\succsim_{\mathbb{Z}}$ satisfies A1-A3 on $\mathcal{I}_{\mathbb{Z}}$. We begin with A1. To see that completeness holds, assume that $I, J \in \mathcal{I}_{\mathbb{Z}}$. There exists $K \in \mathcal{I}$ such that $I + K, J + K \in \mathcal{I}$ and completeness of $\succsim$ on $\mathcal{I}$ implies completeness of $\succsim_{\mathbb{Z}}$ on $\mathcal{I}_{\mathbb{Z}}$. To prove transitivity of $\succsim_{\mathbb{Z}}$, assume that for $I, J, L \in \mathcal{I}_{\mathbb{Z}}$ we have $I \succsim_{\mathbb{Z}} J$ and $J \succsim_{\mathbb{Z}} L$. This means that there are $K, K' \in \mathcal{I}$

such that (i) $I + K \succsim J + K$ and (ii) $J + K' \succsim L + K'$. Applying A2 to $I + K \succsim J + K$ and K' we get $I + K + K' \succsim J + K + K'$, and applying it to $J + K' \succsim L + K'$ and K we have $J + K + K' \succsim L + K + K'$. Transitivity of $\succsim$ on $\mathcal{I}$ yields $I + K + K' \succsim L + K + K'$ and thus $I \succsim_{\mathbb{Z}} L$.

Next we wish to show that $\succsim_{\mathbb{Z}}$ satisfies A2 on $\mathcal{I}_{\mathbb{Z}}$. Let there be given $I, J, K \in \mathcal{I}_{\mathbb{Z}}$, with $I \succsim_{\mathbb{Z}} J$. This means that there exists $K' \in \mathcal{I}$ such that $I + K', J + K'$ are nonnegative (hence in $\mathcal{I}$) and $I + K' \succsim J + K'$. Assume first that $K \in \mathcal{I}$. Then $I + K' + K$ and $J + K' + K$ are nonnegative (and in $\mathcal{I}$). By A2 (of $\succsim$ on $\mathcal{I}$) $I + K' \succsim J + K'$ implies $(I + K') + K \succsim (J + K') + K$ or $(I + K) + K' \succsim (J + K) + K'$ so that $I + K \succsim_{\mathbb{Z}} J + K$. Conversely, if $I + K \succsim_{\mathbb{Z}} J + K$, then for some $K' \in \mathcal{I}$ we have $I + K + K' \succsim J + K + K'$. Then $I \succsim_{\mathbb{Z}} J$ (using $(K + K') \in \mathcal{I}$). Hence we have established that a restriction of A2 holds on $\succsim_{\mathbb{Z}}$: for any nonnegative K we have $I \succsim_{\mathbb{Z}} J$ iff $I + K \succsim_{\mathbb{Z}} J + K$.

For a general (not necessarily nonnegative) $K \in \mathcal{I}_{\mathbb{Z}}$, let $K^+, K^- \in \mathcal{I}$ be such that $K = K^+ - K^-$ (that is, K^+ is equal to K when the latter is positive, and K^- is equal to $-K$ when K is negative). Then $I + K = I + K^+ - K^-$. Denote $I' = I - K^- \in \mathcal{I}_{\mathbb{Z}}$ so that $I = I' + K^-$. Similarly, let $J' = J - K^-$ so that $J = J' + K^-$. By the restricted A2 (applied to the nonnegative $K^- \in \mathcal{I}$), $I' \succsim_{\mathbb{Z}} J'$, iff $I' + K^- \succsim_{\mathbb{Z}} J' + K^-$, that is iff $I \succsim_{\mathbb{Z}} J$. Applying the restricted A2 again to I', J', this time with the non-negative $K^+ \in \mathcal{I}$, we get $I' \succsim_{\mathbb{Z}} J'$ iff $I' + K^+ \succsim_{\mathbb{Z}} J' + K^+$. Thus, $I \succsim_{\mathbb{Z}} J$ iff $I' + K^+ \succsim_{\mathbb{Z}} J' + K^+$. It is left to note that $I' + K^+ = I - K^- + K^+ = I + K$ and similarly $J' + K^+ = J + K$, so that $I \succsim_{\mathbb{Z}} J$ iff $I + K \succsim_{\mathbb{Z}} J + K$.

We now turn to A3. Assume that $I, J, K, L \in \mathcal{I}_{\mathbb{Z}}$, and $K \succ_{\mathbb{Z}} L$. We need to show that, for some $n \geq 0$, we have $I + nK \succ_{\mathbb{Z}} J + nL$. Let $K' \in \mathcal{I}$ be such that $I + K', J + K' \in \mathcal{I}$. As $K \succ_{\mathbb{Z}} L$, there exists $K'' \in \mathcal{I}$ such that $K + K'' \succ L + K''$. Applying A3 (for $\succ$ on $\mathcal{I}$) to $I + K', J + K'$ and the pair $K + K'' \succ L + K''$, we conclude that for some $n \geq 0$ we have

$$I + K' + n(K + K'') \succ J + K' + n(L + K'')$$
or
$$I + nK + (K' + nK'') \succ J + nL + (K' + nK'')$$

Since $K', K'' \in \mathcal{I}$, we also have $K' + nK'' \in \mathcal{I}$. Because $\succsim_{\mathbb{Z}}$ extends $\succsim$, $I + nK + (K' + nK'') \succ J + nL + (K' + nK'')$ is equivalent to $I + nK + (K' + nK'') \succ_{\mathbb{Z}} J + nL + (K' + nK'')$. And, as $\succsim_{\mathbb{Z}}$ is known to satisfy A2 on all of $\mathcal{I}_{\mathbb{Z}}$, this is equivalent to $I + nK \succ_{\mathbb{Z}} J + nL$.

Next, we observe that $\succsim_{\mathbb{Z}}$ is the unique extension of $\succsim$ to $\mathcal{I}_{\mathbb{Z}}$ that satisfies the three axioms. Indeed, should $\succsim'_{\mathbb{Z}}$ be another such extension, for every $I, J \in \mathcal{I}_{\mathbb{Z}}$ there is a $K \in \mathcal{I}$ such that $I + K, J + K \in \mathcal{I}$ and, by A2, we would have $I \succsim'_{\mathbb{Z}} J$ iff $I + K \succsim'_{\mathbb{Z}} J + K$, which (as $\succsim'_{\mathbb{Z}}$ is an extension of $\succsim$) is equivalent to $I + K \succsim J + K$, that is, to $I + K \succsim_{\mathbb{Z}} J + K$. $\square$

Lemma 2 *For all $I, J \in \mathcal{I}_{\mathbb{Z}}$, (i) $I \succsim_{\mathbb{Z}} J$ iff $I - J \succsim_{\mathbb{Z}} 0$; (ii) for all $n > 1$, $I \succsim_{\mathbb{Z}} J$ iff $nI \succsim_{\mathbb{Z}} nJ$.*

Proof: Claim (i) follows from the fact that $\succsim_{\mathbb{Z}}$ satisfies A2 (using $K = -J$). For (ii), note first that, by A2, $I \succsim_{\mathbb{Z}} 0$ iff $kI \succsim_{\mathbb{Z}} (k-1)I$ for all $k \leq n$ (using $K = (k-1)I$). Hence, if $I \succsim_{\mathbb{Z}} 0$, then, by transitivity, $nI \succsim_{\mathbb{Z}} 0$. If $0 \succ_{\mathbb{Z}} I$, the same argument yields $0 \succ_{\mathbb{Z}} nI$. Hence $I \succsim_{\mathbb{Z}} 0$ iff $nI \succsim_{\mathbb{Z}} 0$. Combining this with (i) we conclude that $I \succsim_{\mathbb{Z}} J$ iff $I - J \succsim_{\mathbb{Z}} 0$ iff $nI - nJ \succsim_{\mathbb{Z}} 0$ iff $nI \succsim_{\mathbb{Z}} nJ$. $\square$

We can now make the next step into the rationals.

Lemma 3 *There exists a unique $\succsim_{\mathbb{Q}}$ on $\mathcal{I}_{\mathbb{Q}}$ satisfying A1-A3 and extending $\succsim_{\mathbb{Z}}$ (from $\mathcal{I}_{\mathbb{Z}}$ to $\mathcal{I}_{\mathbb{Q}}$).*

Proof: Let there be given $I, J \in \mathcal{I}_{\mathbb{Q}}$. Define $I \succsim_{\mathbb{Q}} J$ if there exists $n \geq 1$ such that $nI \succsim_{\mathbb{Z}} nJ$ (here and in the sequel, the claim that $\succsim_{\mathbb{Z}}$ holds between two vectors is taken to mean that they are both in $\mathcal{I}_{\mathbb{Z}}$ and that the relation holds between them). Clearly, if $I, J \in \mathcal{I}_{\mathbb{Z}}$ satisfy $I \succsim_{\mathbb{Z}} J$, then $I \succsim_{\mathbb{Q}} J$

(because we may use $n = 1$). Moreover, if $I \succ_{\mathbb{Z}} J$ then we also have $I \succ_{\mathbb{Q}} J$: otherwise, if $J \succsim_{\mathbb{Q}} I$, by Lemma 2 there would be an $n \geq 1$ such that $nJ \succsim_{\mathbb{Z}} nI$ and $J \succsim_{\mathbb{Z}} I$ would follow. Thus, $\succsim_{\mathbb{Q}}$ extends $\succsim_{\mathbb{Z}}$.

We turn to prove that the extended relation $\succsim_{\mathbb{Q}}$ satisfies A1-A3 on $\mathcal{I}_{\mathbb{Q}}$. Start with completeness: given $I, J \in \mathcal{I}_{\mathbb{Q}}$, there exists $n \geq 1$ such that $nI, nJ \in \mathcal{I}_{\mathbb{Z}}$ and completeness of $\succsim_{\mathbb{Z}}$ on $\mathcal{I}_{\mathbb{Z}}$ implies completeness of $\succsim_{\mathbb{Q}}$ on $\mathcal{I}_{\mathbb{Q}}$. To see that transitivity holds as well, assume that $I, J, K \in \mathcal{I}_{\mathbb{Q}}$ with $I \succsim_{\mathbb{Q}} J$ and $J \succsim_{\mathbb{Q}} K$. Let n be such that $nI \succsim_{\mathbb{Z}} nJ$ and let m be such that $mJ \succsim_{\mathbb{Z}} mK$. Applying Lemma 2 to both, we conclude that $nmI \succsim_{\mathbb{Z}} nmJ$ and $nmJ \succsim_{\mathbb{Z}} nmK$ and, by transitivity of $\succsim_{\mathbb{Z}}$, $nmI \succsim_{\mathbb{Z}} nmK$ and $I \succsim_{\mathbb{Q}} K$ is established. Thus $\succsim_{\mathbb{Q}}$ satisfies A1 (on $\mathcal{I}_{\mathbb{Q}}$ in its entirety).

Next consider A2. Let there be given $I, J, K \in \mathcal{I}_{\mathbb{Q}}$, and assume first that $I \succsim_{\mathbb{Q}} J$, where we need to show that $I + K \succsim_{\mathbb{Q}} J + K$. Since $I \succsim_{\mathbb{Q}} J$, there exists $n \geq 1$ such that $nI \succsim_{\mathbb{Z}} nJ$. Because $K \in \mathcal{I}_{\mathbb{Q}}$, for some $m \geq 1$ we have $mK \in \mathcal{I}_{\mathbb{Z}}$. In light of Lemma 2, we also have $nmI \succsim_{\mathbb{Z}} nmJ$. Apply A2 (on $\succsim_{\mathbb{Z}}$ on $\mathcal{I}_{\mathbb{Z}}$) for $nmK \in \mathcal{I}_{\mathbb{Z}}$ to obtain $nmI + nmK \succsim_{\mathbb{Z}} nmJ + nmK$ or $nm\,(I + K) \succsim_{\mathbb{Z}} nm\,(J + K)$ – that is, $I + K \succsim_{\mathbb{Q}} J + K$. To see the converse, if $I + K \succsim_{\mathbb{Q}} J + K$ we may define $I' = I + K$, $J' = J + K$ and $K' = -K$ and use the previous proof.

We now turn to A3. Assume that $I, J, K, L \in \mathcal{I}_{\mathbb{Q}}$, and $K \succ_{\mathbb{Q}} L$. We need to show that, for some $n \geq 0$, we have $I + nK \succ_{\mathbb{Q}} J + nL$. Let $m \geq 1$ satisfy $mJ, mI \in \mathcal{I}_{\mathbb{Z}}$. By $K \succ_{\mathbb{Q}} L$ we know that for some $m' \geq 1$, $m'K \succ_{\mathbb{Z}} m'L$. Clearly, $mm'I, mm'J \in \mathcal{I}_{\mathbb{Z}}$. By Lemma 2, from $m'K \succ_{\mathbb{Z}} m'L$ we conclude that $mm'K \succ_{\mathbb{Z}} mm'L$. Applying A3 (on $\succsim_{\mathbb{Z}}$ on $\mathcal{I}_{\mathbb{Z}}$) to $mm'I, mm'J, mm'K, mm'L$ we infer that there exists $n \geq 0$ such that

$$mm'I + nmm'K \succ_{\mathbb{Z}} mm'J + nmm'L$$

or

$$mm'\,(I + nK) \succ_{\mathbb{Z}} mm'\,(J + nL)$$

which, by definition, means that $I + nK \succ_{\mathbb{Q}} J + nL$.

Finally, we observe that the extension $\succsim_{\mathbb{Q}}$ is unique: should another relation, $\succsim'_{\mathbb{Q}}$ on $\mathcal{I}_{\mathbb{Q}}$ also extend $\succsim_{\mathbb{Z}}$ and satisfy A1-A3 on $\mathcal{I}_{\mathbb{Q}}$, for every $I, J \in \mathcal{I}_{\mathbb{Q}}$

we can find $n \geq 1$ such that $nI, nJ \in \mathcal{I}_{\mathbb{Z}}$ and then $I \succsim'_{\mathbb{Q}} J$ iff $nI \succsim_{\mathbb{Z}} nJ$, iff $I \succsim_{\mathbb{Q}} J$. $\square$

Lemma 4 *For all $I, J \in \mathcal{I}_{\mathbb{Q}}$, (i) $I \succsim_{\mathbb{Q}} J \iff I - J \succsim_{\mathbb{Q}} 0$; (ii) for all $n > 1$, $I \succsim_{\mathbb{Q}} J \iff nI \succsim_{\mathbb{Q}} nJ$.*

Proof: Identical to the proof of Lemma 2. $\square$

It is worth mentioning explicitly the following:

Lemma 5 *For all $I, J \in \mathcal{I}_{\mathbb{Q}}$, and every rational number $\alpha > 0$, $I \succsim_{\mathbb{Q}} J$ iff $\alpha I \succsim_{\mathbb{Q}} \alpha J$ (and thus $I \succ_{\mathbb{Q}} J$ iff $\alpha I \succ_{\mathbb{Q}} \alpha J$).*

Proof: Let there be given $I, J \in \mathcal{I}_{\mathbb{Q}}$ and $\alpha = \frac{k}{n}$ for $k, n > 0$. By Lemma 4 applied to I, J and k, we know that $I \succsim_{\mathbb{Q}} J$ iff $kI \succsim_{\mathbb{Q}} kJ$. Applying it to $\alpha I, \alpha J$ and n, we also know that $\alpha I \succsim_{\mathbb{Q}} \alpha J$ iff $n\alpha I \succsim_{\mathbb{Q}} n\alpha J$. However, $kI = n\alpha I$ and $kJ = n\alpha J$. It follows that $I \succsim_{\mathbb{Q}} J$ iff $\alpha I \succsim_{\mathbb{Q}} \alpha J$. $\square$

We turn to the definition of the function u such that $\sum_{p \in P} I(p)\, u(p) \geq \sum_{p \in P} J(p)\, u(p)$ iff $I \succsim J$. In the trivial case where $I \sim J$ for all $I, J \in \mathcal{I}$, we can select $u(\cdot) \equiv 0$. Clearly, it is the only function that represents $\succsim$ in this case. We therefore focus on the more interesting case in which $I^* \succ J^*$ for some $I^*, J^* \in \mathcal{I}$. Let $P^* = supp(I^*) \cup supp(J^*)$ and consider a finite $P_0 \subset P$ such that $P^* \subset P_0$. We will define u for such a subset P_0 and then show how these definitions can be "patched" together to a definition of u on all of P.

Given such a $P_0 \subset P$, consider the vectors $\mathcal{I}_0$ whose support is in P_0. Define $\mathcal{I}_{\mathbb{Q}_0}$ accordingly (as the vectors in $\mathcal{I}_{\mathbb{Q}}$ that vanish outside of P_0). Next, define

$$
\begin{aligned}
A &= \{ I \in \mathcal{I}_{\mathbb{Q}_0} \mid I \succ_{\mathbb{Q}} 0 \} \\
B &= \{ I \in \mathcal{I}_{\mathbb{Q}_0} \mid 0 \succ_{\mathbb{Q}} I \} \\
E &= \{ I \in \mathcal{I}_{\mathbb{Q}_0} \mid I \sim_{\mathbb{Q}} 0 \}
\end{aligned}
$$

We consider A, B, E as subsets of the finite-dimensional Euclidean space $\mathbb{R}^{|P_0|}$, and observe the following:

1. A, B, E are pairwise disjoint – by definition of $\sim_{\mathbb{Q}}, \succ_{\mathbb{Q}}$;

2. $A \cup B \cup E = \mathcal{I}_{\mathbb{Q}_0}$ – by completeness;

3. $A = -B$ – by A2 applied to $\succsim_{\mathbb{Q}}$;

4. $0 \in E$ – because $\succsim_{\mathbb{Q}}$ is reflexive;

5. For all $I, J \in \mathcal{I}_{\mathbb{Q}_0}$, $I \succ_{\mathbb{Q}} J \iff I - J \in A$, $J \succ_{\mathbb{Q}} I \iff I - J \in B$, and $I \sim_{\mathbb{Q}} J \iff I - J \in E$ – by Lemma 4 and A2; and

6. $A, B \neq \varnothing$ – because we have $I^* \succ_{\mathbb{Q}} J^*$.

We now wish to show that A, B, and E are convex over the rationals.

Lemma 6 *Suppose that $I, J \in A(B, E)$ and $\alpha \in [0,1] \cap \mathbb{Q}$. Then $\alpha I + (1-\alpha) J \in A(B, E)$.*

Proof: Let there be given $I, J \in A$ and $\alpha I + (1-\alpha) J$ for $\alpha \in [0,1] \cap \mathbb{Q}$. Assume wlog that $\alpha \in (0,1)$. Since $I, J \succ_{\mathbb{Q}} 0$, we can use Lemma 5 to conclude that $\alpha I, (1-\alpha) J \succ_{\mathbb{Q}} 0$. From $\alpha I \succ_{\mathbb{Q}} 0$ A2 implies (for $K = (1-\alpha) J$) that $\alpha I + (1-\alpha) J \succ_{\mathbb{Q}} (1-\alpha) J$. By transitivity, we thus have $\alpha I + (1-\alpha) J \succ_{\mathbb{Q}} 0$, that is, $\alpha I + (1-\alpha) J \in A$.

The proof for B is identical, and its convexity can also be derived from the fact that $B = -A$. Finally, the argument for the convexity of E is very similar, using the weak-preference part of A2 in both directions. □

In the following, for $I \in \mathbb{R}^{|P_0|}$ and $\varepsilon > 0$, $N_\varepsilon(I)$ denotes the ε-neighborhood of I in $\mathbb{R}^{|P_0|}$.

Lemma 7 *Let $I \in A(B)$. There exists $\varepsilon > 0$ such that for every $J \in N_\varepsilon(I) \cap \mathcal{I}_{\mathbb{Q}_0}$, we have $J \in A(B)$.*

Proof: Let there be given $I \in A$. For $1 \leq j \leq |P_0|$, let e_j be the j-th unit vector in $\mathbb{R}^{|P_0|}$. We wish to find an $\varepsilon_j > 0$ such that $[I - \varepsilon_j e_j, I + \varepsilon_j e_j] \cap \mathcal{I}_{\mathbb{Q}_0} \subset A$, that is, a positive-size interval around I in the direction of the j-th dimension, so that all rational points on it are in A. To this end, we will show that (i) there exists $\varepsilon_j^+ > 0$ such that $\left[I, I + \varepsilon_j^+ e_j\right] \cap \mathcal{I}_{\mathbb{Q}_0} \subset A$, and then (ii) there exists $\varepsilon_j^- > 0$ such that $\left[I - \varepsilon_j^- e_j, I\right] \cap \mathcal{I}_{\mathbb{Q}_0} \subset A$ – and then define $\varepsilon_j = \min\left(\varepsilon_j^+, \varepsilon_j^-\right)$. Starting with ε_j^+, consider $K^+ = I + e_j \in \mathcal{I}_{\mathbb{Q}_0}$. By A3 there exists $n \geq 1$ such that $K^+ + nI \succ_{\mathbb{Q}} 0$. Hence by Lemma 4), $\frac{1}{n+1}\left[K^+ + nI\right] \succ_{\mathbb{Q}} 0$ and we can take $\varepsilon_j^+ = \frac{1}{n+1}$. Clearly, the argument for the existence of a suitable ε_j^- is symmetric.

Let $\bar{\varepsilon} = \min_j \varepsilon_j$ and set $\varepsilon = \frac{\bar{\varepsilon}}{\sqrt{|P_0|}} > 0$. It follows that the ε-ball around I is included in the convex hull of $\{I - \bar{\varepsilon} e_j, I + \bar{\varepsilon} e_j\}_{j \leq |P_0|}$. Consider a rational point $J \in N_\varepsilon(I) \cap \mathcal{I}_{\mathbb{Q}_0}$. By Carathéodory's Theorem, there are $l \leq |P_0| + 1$ points $\{I_r\}_{r \leq l}$ in $\{I - \bar{\varepsilon} e_j, I + \bar{\varepsilon} e_j\}_{j \leq |P_0|}$ and nonnegative numbers $\{\alpha_r\}_{r \leq l}$ adding up to 1 such that $J = \sum_{r \leq l} \alpha_r I_r$, and, furthermore, these $\{\alpha_r\}_{r \leq l}$ are unique (for the given set $\{I_r\}_{r \leq l}$). This implies that $\{\alpha_r\}_{r \leq l}$ are rational (as the unique solution to a system of equations with rational numbers). It follows, from iterated applications of of Lemma 6, that $J = \sum_{r \leq l} \alpha_r I_r \in A$.

The argument for B is symmetric. □

Lemma 8 *For every $I \in E$ there are arbitrarily close points $J^+ \in A$ and $J^- \in B$.*

Proof: Consider $I \in E$ and let $J \in A$. Then, by Lemma 4, A2, and transitivity, for every $n \geq 1$, $nI + J \in A$, and this means (by Lemma 4 again) that $\frac{1}{n+1}\left[nI + J\right] \in A$ – for every $n \geq 0$. The argument for B is obviously symmetric. □

Lemma 9 *There exists $u : P_0 \to \mathbb{R}$ such that, for all $I, J \in \mathcal{I}_{\mathbb{Q}_0}$*

$$
\begin{array}{rcl}
I & \succsim & {}_{\mathbb{Q}_0} J \\
 & \textit{iff} & \\
\sum_{p \in P_0} I(p)\, u(p) & \geq & \sum_{p \in P_0} J(p)\, u(p)
\end{array}
$$

Further, u is unique up to multiplication by a positive number.

Proof: We know that A and B are closed under rational convex combinations. Consider $A^0 = int\left(conv\left(A\right)\right)$ and $B^0 = int\left(conv\left(B\right)\right)$, where *conv* refers to the convex hull in $\mathbb{R}^{|P_0|}$ (including not-necessarily rational points, and not-necessarily rational convex combinations). By Lemma 7 the sets $A^0, B^0 \subset \mathbb{R}^{|P_0|}$, which are open and convex, are non-empty (and full-dimensional). By Lemma 7 we also know that $A \subset A^0$ and $B \subset B^0$. Moreover, for any $I \in A^0(B^0)$ there exists $\varepsilon > 0$ such that $N_\varepsilon(I) \cap \mathcal{I}_{\mathbb{Q}_0} \subset A(B)$. In particular, this means that $A^0 \cap B^0 = \varnothing$: if there were $x \in A^0 \cap B^0$, we would be able to find two neighborhoods thereof, such that in one all rational points have to be in A and in the other – in B, in contradiction to $A \cap B = \varnothing$.

Thus we can apply a separating hyperplane theorem to conclude that there exists $u : P_0 \to \mathbb{R}$ and a number $c \in \mathbb{R}$ such that

$$\begin{array}{llll} I \in & A^0 & \Longrightarrow & I \cdot u > c \qquad \text{(i)} \\ I \in & B^0 & \Longrightarrow & I \cdot u < c \qquad \text{(ii)} \end{array}$$

We first note that the implications in (i) and in (ii) can also be reversed. To see this, assume that $I \cdot u > c$. Choose $\varepsilon > 0$ such that for every $J \in N_\varepsilon(I)$, we have $J \cdot u > c$. We argue that, if $J \in \mathcal{I}_{\mathbb{Q}_0}$, we have $J \in A$. Indeed, if $J \in B$ then $J \in B^0$ and (ii) would imply $J \cdot u < c$. If, however, $J \in E$, then, by Lemma 8, there are points $J' \in B$ arbitrarily close to J, so that we can pick $J' \in N_\varepsilon(I) \cap B$. This would contradict (ii) again. Thus all rational points in $N_\varepsilon(I)$ are in A and $I \in A^0$ follows.

The argument is symmetric for B (and the converse of (ii)).

Next, observe that for $I \in E$ we have to have $I \cdot u = c$, because an inequality would imply that I is in A or in B.

Finally we observe that $c = 0$ because $0 \in E$.

We conclude that for all $I \in \mathcal{I}_{\mathbb{Q}}$

$$\begin{array}{lll} I \in A & \Longleftrightarrow & I \cdot u > 0 \\ I \in E & \Longleftrightarrow & I \cdot u = 0 \\ I \in B & \Longleftrightarrow & I \cdot u < 0 \end{array}$$

which means that, for all $I, J \in \mathcal{I}_{\mathbb{Q}_0}$, $I \succsim_{\mathbb{Q}_0} J$ $(I \succ_{\mathbb{Q}_0} J)$ iff $I - J \in A \cup E$ $(I - J \in A)$ iff $(I - J) \cdot u \geq 0$ (> 0) – and u is unique up to multiplication by a positive constant $\lambda > 0$. $\square$

To conclude the proof of the theorem, we now define u for all p. We apply Lemma 9 and fix one $u_0 : P_0 \to \mathbb{R}$ that represents $\succsim_{\mathbb{Q}_0}$ on $\mathcal{I}_{\mathbb{Q}_0}$. For every $p \notin P_0$, we consider $P_p = P \cup \{p\}$ and the corresponding $\mathcal{I}_{\mathbb{Q}_p}$. There exists a unique function u_p that extends u_0 to P_p and represents $\succsim_{\mathbb{Q}_p}$. Let $u(p) = u_p(p)$. Thus u is defined for all of P.

To see that u so defined represents $\succsim_{\mathbb{Q}}$ on all of $\mathcal{I}_{\mathbb{Q}}$, consider $K, L \in \mathcal{I}_{\mathbb{Q}}$. Limit attention to $P' = P^* \cup supp(K) \cup supp(L)$ (where P^* was the subset used for the construction above, over which preferences are non-trivial). By the same arguments, there exists a $u' : P' \to \mathbb{R}$ that represents $\succsim_{\mathbb{Q}'}$. It thus also represents $\succsim_{\mathbb{Q}_0}$ on $\mathcal{I}_{\mathbb{Q}_0}$ and $\succsim_{\mathbb{Q}_p}$ on $\mathcal{I}_{\mathbb{Q}_p}$ for every $p \in P'$. Hence it has to equal u up to multiplication by a positive constant.

7 Bibliography

Adler, Matthew D. (2012), *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis.* Oxford University Press.

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

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

Blackorby, Charles, Walter Bossert, and David Donaldson (1995), "Intertemporal Population Ethics: Critical-Level Utilitarian Principles", *Econometrica*, **65**: 1303-20.

Blackorby, Charles, Walter Bossert, and David Donaldson (1997), "Critical-Level Utilitarianism and the Population Ethics Dilemma", *Economics and Philosophy*, **13**: 197-230.

Blackorby, Charles and David Donaldson (1984), "Social Criteria for Evaluation Population Change", *Journal of Public Economics*, **25**: 13-33.

Boonin, David (2014), *The Non-Identity Problem and the Ethics of Future People*, Oxford University Press.

Bossert, Walter (1990), "Maximin welfare orderings with variable population size", *Social Choice and Welfare*, **7**: 39-45.

Broome, John (2004), *Weighing Lives*, Oxford University Press.

de Finetti, Bruno (1931), "Sul Significato Soggettivo della Probabilità ", *Fundamenta Mathematicae*, **17**: 298-329.

de Finetti, Bruno (1937), "La Prevision: Ses Lois Logiques, Ses Sources Subjectives", *Annales de l'Institut Henri Poincaré* , **7**: 1-68.

Carlson, Erik (1998), "Mere Addition and the Two Trilemmas of Population Ethics", *Economics & Philosophy*, **14**: 283-306.

Frick, Johann (2020), "Conditional Reasons and the Procreation Asymmetry", *Philosophical Perspectives*, **34**: 53-87.

Fryxell, Loren K. (2024), "XU" (Experienced Utility), working paper.

Gilboa, Itzhak (2009), *Theory of Decision under Uncertainty*. Econometric Society Monograph Series Vol. No. 45. Cambridge: Cambridge University Press.

Gilboa, Itzhak and David Schmeidler (2001), "A Cognitive Model of Individual Well-Being", *Social Choice and Welfare*, **12**: 269-288.

Hammond, Peter J. (1976), "Equity, Arrow's Conditions, and Rawls' Difference Principle", *Econometrica,* **44**: 793-804.

Harsanyi, John C. (1955), "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy*, **63**: 309-321; reprinted in Phelps (1973) and in Harsanyi (1976).

Huemer, Michael (2008), "In Defence of Repugnance", *Mind*, **117**(468): 899-933.

Hurka, Thomas (1983), "Value and Population Size", *Ethics*, **93**: 496-507.

Luce, R. Duncan and David H. Krantz (1971), “Conditional Expected Utility”, *Econometrica*, **39**: 253-271.

Narveson, Jan (1973), “Moral Problems of Population”, *The Monist*, **57**: 62-86.

Parfit, Derek (1976), “On Doing the Best for Our Children”, in: M. Bayles. ed.. *Ethics and Population* (Schenkman, Cambridge).

Parfit, Derek (1982), “Future generations: Further problems”, *Philosophy and Public Affairs*, **11**: 113-172.

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

Parfit, Derek (1997), “Equality and Priority”, *Ratio,* **10**: 202-221.

Persson Ingmar (2004), “The Root of the Repugnant Conclusion and its Rebuttal”, in J. Ryberg and T. Tännsjö (eds.) *The Repugnant Conclusion*, Library of Ethics and Applied Philosophy: 187-200.

Quine, Willard Van Orman (1962), *Methods of Logic*, London: Routledge & Kegan Paul.

Rachels, Stuart (2004), “Repugnance or Intransitivity: A Repugnant but Forced Choice”, in J. Ryberg and T. Tännsjö (eds.) *The Repugnant Conclusion*, Library of Ethics and Applied Philosophy: 163-186.

Roberts, Melinda A. (2011), “The Asymmetry: A Solution ”, *Theoria,* **77**: 333-367.

Savage, Leonard J. (1954), *The Foundations of Statistics.* New York: John Wiley and Sons. (Second addition in 1972, Dover)

Sen, Amartya K. (1977), “On weights and measures: Informational constraints in social welfare analysis ”, *Econometrica*, **45**: 1539-1572.

Sikora, Richard I. (1975), “Utilitarianism: The Classical Principle and the Average Principle ”. *Canadian Journal of Philosophy,* **5**: 409-419.

Sikora, R. I. (1978). “Is It Wrong to Prevent the Existence of Future Generations?” In R. I. Sikora & B. Barry (Eds.), *Obligations to Future Generations* (pp. 112-166). Temple University Press.

Strawson, Sir Peter Frederick (1950), “On Referring ”, *Mind*, **59**: 320-344.

Tännsjö , Torbörn (2002), “Why We Ought to Accept the Repugnant Conclusion”, *Utilitas*, **14**: 339-350.

Temkin, Larry S. (2012), *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning.* Oxford University Press.

Temkin, Larry S. (1987), “Intransitivity and the Mere Addition Paradox”, *Philosophy and Public Affairs*, **16**: 138-187.

von Neumann, John and Oskar Morgenstern (1944, 1947), *Theory of Games and Economic Behavior*. Princeton, N.J.: Princeton University Press.

Zuber, Stéphane and Geir B. Asheim (2012), “Justifying Social Discounting: The Rank-Discounted Utilitarian Approach”, *Journal of Economic Theory,* **147**: 1572-1601.