CESifo WORKING PAPERS

6613
2017

August 2017

Revealed Relative Utilitarianism

Tilman Börgers, Yan-Min Choo

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 6613*
Category 2: Public Choice

Revealed Relative Utilitarianism

Abstract

We consider the aggregation of individual agents' von Neumann- Morgenstern preferences over lotteries into a social planner's von Neumann-Morgenstern preference. We start from Harsanyi's [18] axiomatization of utilitarianism, and ask under which conditions a social preference order that satisfies Harsanyi's axiom uniquely reveals the planner's marginal rates of substitution between the probabilities of any two agents' most preferred alternatives, assuming that any increase/decrease in the probability of each agent's most preferred alternative is accompanied by an equally sized decrease/increase in that agent's least preferred alternative. We then introduce three axioms for these revealed marginal rates of substitution. The only welfare function that satisfies these three axioms is the relative utilitarian welfare function. This welfare function, that was introduced in Dhillon [9] and Dhillon and Mertens [11], normalizes all agents' utility functions so that the lowest value is 0 and the highest value is 1, and then adds up the utility functions. Our three axioms are closely related to axioms that Dhillon and Mertens used to axiomatize relative utilitarianism. We simplify the axioms, provide a much simpler and more transparent derivation of the main result, and re-interpret the axioms as revealed preference axioms.

JEL-Codes: D600.

Tilman Börgers
Department of Economics
University of Michigan
USA - 48109 Ann Arbor Michigan
tborgers@umich.edu

Yan-Min Choo
Singapore
ychoo@umich.edu

August 10, 2017
We are grateful to Lars Ehlers, Yusufcan Masatlioglu, Philippe Mongin, John Weymark, and three referees for comments and discussions. We owe special thanks to Jim Belk who suggested to us the idea on which the proof of Proposition 3 is built.

1 Introduction

The Revealed Preference Approach to Welfare Judgments

Welfare judgments are ubiquitous in economics. One of the most prominent welfare functions is the utilitarian welfare function, according to which welfare equals a weighted sum of individuals' utilities. There are several approaches to understanding and justifying utilitarianism in economics. For this paper the difference between two particular such approaches is important. The first approach is to assume that, when making welfare judgments, we are *given* not only individuals' ordinal preferences over alternatives, but also some numerical utility information the meaning of which goes beyond the mere representation of ordinal preferences. For example, utility information might reflect the strength of individuals' preferences, or it might indicate how utilities of different people compare to each other. Mathematically speaking, in this approach the argument of the welfare function are utility functions rather than preferences. The literature on this approach is vast. Examples of theorems that axiomatize utilitarianism in the framework of this approach are results due to d'Aspremont and Gevers [7], Maskin [23], and Mongin [24].[1]

The second approach to the justification of utilitarianism only uses individual preferences as input into the social welfare assessment. This approach originates with Harsanyi [18]. Harsanyi assumed every individual in society to have von Neumann-Morgenstern preferences regarding all lotteries over a given set of alternatives. Society's preference over lotteries is also assumed to satisfy the von Neumann-Morgenstern axioms. In this setting, utilitarianism means that any Bernoulli utility function representing society's preferences is a weighted sum of Bernoulli utility functions representing the individuals' preferences.[2] Harsanyi showed that utilitarianism in this sense is implied by a simple "indifference axiom:" if all individuals are indifferent between two lotteries, then so should society be.[3] Importantly, the indifference axiom is a single profile axiom, that is, it can be

[1]More references can be found, for example, in the survey by d'Aspremont and Gevers [8].

[2]It has been disputed that Harsanyi's [18] theorem is "really" about utilitarianism because it treats utility functions only as representations of ordinal preferences rather than as primitive concepts (see Weymark's [27] review of the Harsanyi-Sen debate). For simplicity, we shall ignore this issue in this paper.

[3]Note that this argument is different from the justification of utilitarianism as the

checked for every profile of preferences separately. The indifference axiom implies no restrictions across preference profiles. Also, even if we fix some arbitrary choice of Bernoulli utility functions to represent individuals' preferences, the indifference axiom does not pin down the weights assigned to different individuals' utility functions. The weights may even be negative, although this is easily ruled out by strengthening the indifference axiom to a Pareto axiom.[4]

The Harsanyi weights, together with the Bernoulli utility function chosen to represent each individual's preference, determine the "importance," informally speaking, of this individual's preferences for social preferences.[5] Because Harsanyi's single profile approach does not imply any restrictions across preference profiles, the importance of an individual for social preferences may depend on the preference profile that is considered. Thus, for example, an individual who ranks alternative a over alternative b may be considered to display "bad taste" and therefore may be almost irrelevant to social preferences (unless all individuals rank a over b). In this way, the welfare function may incorporate a form of "paternalism." Also an individual whose preferences deviate particularly strongly from those of all others may be given particularly great importance, which might be justified as a form of "minority protection." All of this is allowed by Harsanyi's theorem.

"Relative utilitarianism" is one way to go further, and to come up with a single welfare definition. Relative utilitarianism defines welfare to be the sum of agents' Bernoulli utilities where all agents have the same weight, and where agents' Bernoulli utility functions are normalized so that each individual's utility function assigns utility 0 to this individual's least preferred alternative and utility 1 to this individual's most preferred alternative. The

rational choice criterion for a fictitious observer choosing behind the veil of ignorance, that is, not knowing which individual she herself will be in society. This justification was proposed by Harsanyi in [17]. In the setting of Harsanyi's [17] paper, an axiomatization of the welfare definition that we discuss in this paper, "relative utilitarianism," was provided by Karni [22].

[4] For a modern statement and proof of Harsanyi's theorem, and of versions of the theorem with the indifference axiom strengthened to a variety of Pareto axioms, see Weymark [28].

[5] To obtain an analytically clear formalization of the "importance" of an individual's preference for social preferences it is useful to use language that refers to preferences only rather than to their numerical representations. We do this later, by introducing the concept of "marginal rates of substitution." For the purposes of this paragraph, we shall use the, in this context inevitably vague, term "importance."

subject of this paper is the axiomatic basis of relative utilitarianism.

To develop this axiomatic perspective we adopt the "revealed preference" approach to welfare judgments.[6] Suppose we observed sufficiently many choices of each member of society to infer their preferences, and also sufficiently many choices of a "social planner" to infer the planner's preferences. Suppose all these preferences satisfied von Neumann and Morgenstern's axioms. Assume also that we made observations not just for one profile of individuals' preferences but for many. And finally, imagine that we knew that the social planner, when choosing, knew the profile of individuals' preferences, and that his choices satisfy a strengthened version of Harsanyi's indifference axiom, namely a Pareto axiom. Then the social planner's choices reveal, in a sense that we shall make precise, for every preference profile how much weight the planner assigns to any particular individual in comparison to any other individual.

More precisely, we shall introduce a concept of the social planner's "marginal rates of substitution" between the probabilities of any individual i and j's most preferred alternatives, assuming that any increase/decrease in the probability of some agent's most preferred alternative is accompanied by an equally sized decrease/increase in the agent's least preferred alternative. We shall explain how these marginal rates of substitution are revealed by the planner's choices. We shall show that three axioms about the social planner's choices imply that these marginal rates of substitution have to be equal to 1 for all preference profiles, and we shall infer from this result that the planner's choices must reveal a relative utilitarian welfare function.

Before we elaborate on the axioms, we emphasize that our revealed preference approach is positive, not normative. In particular, the importance of different individuals' preferences for the planner's preference, i.e. how the planner compares different agents' utilities, is revealed. We don't ask

[6] This approach has also been adopted in the literature in which utility functions are arguments of the social welfare function (Mongin [24]). In that approach, the preferences that are revealed are only the planner's preferences. It is not discussed who individuals' utility functions are revealed. In our context, both the individuals' and the planner's preferences can, in principle, be revealed through choices. Binmore [4, Chapter 4] advocates the interpretation of Harsanyi's utilitarianism theorem as a theorem about revealed preferences, which is the interpretation we use, but does not consider the multi-profile extension of Harsanyi's theory. Fishburn ([12], [13]) has developed what can be interpreted as a revealed preference approach to utilitarianism in a setting without lotteries.

how interpersonal comparisons "should" be made. Therefore, the normative plausibility of our axioms is irrelevant. Their plausibility as a description of real world social decision making is what matters. However, our purpose in this paper is also not to evaluate the positive plausibility of the axioms. Our contribution is merely to show that these axioms are equivalent to revealed relative utilitarianism.

Axioms for Revealed Relative Utilitarianism

As we mentioned above, the key concept in our axiomatization of relative utilitarianism are the "marginal rates of substitutions" of the welfare function. The "marginal rates of substitution" are defined for a given profile of individuals' von Neumann-Morgenstern preferences and are derived from the social planner's von Neumann-Morgenstern preference that corresponds to this profile. We assume that the social planner's preference satisfies a Pareto axiom. The marginal rate of substitution between agents i and j is then the answer to the following question: To keep the social planner indifferent, how much probability of agent i's most preferred alternative must be shifted to his least preferred alternative, if "one small unit of probability" is shifted from agent j's least preferred alternative to his most preferred alternative? In this question we assume implicitly that the shift in probabilities from agent i's most preferred alternative to his least preferred alternatives does not affect the utility of agents other than i, and we make the same implicit assumption for agent j. Using an idea due to Weymark [28] we show that, when the preference profiles satisfies a condition that is called the "Independent Prospects" condition, this is a well-defined thought experiment for all i and j, and for every pair of i and j there is a unique number that is the answer to our question.

We then consider a social planner who assigns to many profiles of individuals' preferences a social preference that satisfies the Pareto axiom. Whenever the profile of individuals' preferences satisfies the Independent Prospects condition, then the planner's preference reveals a complete set of pairwise marginal rates of substitutions. The axioms that we study in this paper address how these marginal rates of substitution change as individuals' preferences change. When the social preference can be represented by the relative utilitarian welfare function, then the marginal rates of substitution equal 1, regardless of individuals' preferences. The objective of our analysis is to derive this as a conclusion from more elementary axioms.

The key axiom is a separability axiom. It requires that the marginal rate of substitution between the probabilities of agent i's and agent j's most preferred alternatives (at the expense of their least preferred alternatives) only depends on agent i and agent j's preferences, not on other agents' preferences. Any preference that satisfies (i) a Pareto axiom and (ii) the separability axiom can be represented as the sum of 0-1 normalized Bernoulli utilities such that each agent's weight depends only on that agent's preference, and not on the other agents' preferences. Separability thus rules out in particular that an agent's weight depends on the comparison between her preference and other agents' preferences, as in the "minority protection" example that we offered above.

We then add the "invariance axiom." This axiom requires that the marginal rates of substitution do not change when a change in agents' preferences concerns only alternatives which all agents regard as equivalent to lotteries over the other alternatives, and which are thus "redundant." What changes in these cases is only agents' views of *which* lotteries the redundant alternatives are equivalent to. Such a change is required by our axiom not to affect the marginal rates of substitution. A short argument that is conceptually and mathematically not very deep shows that the invariance axiom has a surprisingly strong implication if the domain of preferences that is considered is sufficiently rich. It then implies that in the representation of social welfare as the sum of 0-1 normalized von Neumann-Morgenstern utilities each agent's weight must be constant, and not vary with the preference profile at all. Thus, in particular, the invariance axiom rules out the paternalism example that we offered above.

The final step of the argument is simple. We add an anonymity axiom, which requires that all agents are treated symmetrically. We then conclude that all agents' weights have to be equal, and thus that social welfare has to be relative utilitarian.

It is of technical importance for our analysis that we restrict our argument to a sub-domain of the space of all profiles of von Neumann-Morgenstern preferences. For example, we only consider preference profiles that satisfy the "Independent Prospects" condition mentioned earlier. There will be other constraints on the sub-domain that we are considering. In fact, this sub-domain is not explicitly constructed in the paper. Instead we list in the paper all properties of the sub-domain that we make use of, and we construct the sub-domain explicitly in the appendix. The sub-domain is in a natural sense "dense" in the universal domain, if we exclude from the universal domain all profiles in which some individual is completely indifferent between all alternatives. The construction of the sub-domain is somewhat artificial. Whether there is a natural axiomatization of relative utilitarianism on the domain of all profiles of von Neumann-Morgenstern preferences is an open question.

We will be able to construct the sub-domain of preferences for which our results holds only under restrictive assumptions regarding the number of agents and alternatives. Specifically, we require that there are at least three agents, and that the number of alternatives is more than six times the number of agents. The paper will make clear that the assumptions that there are more alternatives than agents, and that there are at least three agents, are made for transparent conceptual reasons. The final assumption that the number of alternatives exceeds the number of agents by a factor of more than six is made for purely mathematical reasons which the appendix clarifies. That there are many more alternatives than agents seems not unreasonable: there are probably many more possible income tax codes than there are citizens of the United States.

Relation with Dhillon and Mertens's Axiomatization of Relative Utilitarianism

Our axiomatization of relative utilitarianism is closely related to an axiomatization of relative utilitarianism due to Dhillon [9]. Our three axioms are in fact very similar to the three axioms that Dhillon uses.[7] Dhillon derives the separability axiom from a more complicated axiom, the "extended Pareto axiom," that can only be formulated in a richer framework than the one we consider here. This richer framework considers the aggregation of preferences in groups of different size. It is easy to see that the extended Pareto axiom implies the separability axiom (see our discussion following Definition 6). Using separability directly makes our paper much simpler than it would otherwise be.

Unlike Dhillon, we adopt a revealed preference approach to social welfare. In particular, we use throughout the new concept of revealed marginal rates of substitution of a social welfare function. The introduction of this concept is another factor that allows us to provide much simpler and more

[7] Dhillon introduces a fourth axiom, "neutrality," but never uses it.

transparent proofs than Dhillon offered. Our approach of considering, as we mentioned earlier, only a sub-domain of the full domain of preferences also contributes to the simplification that we achieve in this paper. Dhillon works with the full domain, and, remarkably, does not use any continuity axiom.[8]

A paper that is closely related to Dhillon [9] is Dhillon and Mertens [11]. They offer an axiomatization of relative utilitarianism that differs from Dhillon's by replacing the extended Pareto axiom by a very weak "monotonicity" axiom, and then requiring continuity of the social welfare function.[9] If continuity is not assumed, Dhillon and Mertens' axioms allow, for example, the weights of any one individual to depend on the number of other individuals with the same preferences.[10]

Relation with Harsanyi's Work

Our interpretation of welfare theory as a theory of the revealed preferences of a social planner echoes language used by Harsanyi when summarized the conclusion of his 1955 paper thus: "In the same way as ... it has been shown that a rational man ... must act *as if* he ascribed numerical subjective probabilities to all alternative hypotheses ... - so in welfare economics we have also found that a rational man ... must likewise act *as if* he made quantitative interpersonal comparisons of utility..." [18, p. 321].

Harsanyi [18] discussed in detail an earlier paper by Fleming [15] that also provided axiomatic foundations for weighted utilitarianism, interpreting his own work as reaching the same conclusion as Fleming's but with weaker axioms. Harsanyi thought that the main difference between his and Fleming' s framework that made it possible to drop axioms was that he considered lotteries as alternatives, while Fleming did not, and then could

[8]Unfortunately, Dhillon [9] contains errors that affect both the statement of the main result and its proof. We elaborate on this in Börgers and Choo [5]. We do not know whether statement and proof of the main result in Dhillon [9] can be repaired.

[9]The continuity notion that Dhillon and Mertens use is intricate. It needs to be, because with a more simple notion of continuity, Chichilnisky's [6] impossibility result that we mention below would apply.

[10]See the example on page 485 in Dhillon and Mertens [11], which they use to illustrate that the continuity axiom cannot be dropped from the theorem. This example also disproves the assertion of Dhillon and Mertens on page 483 of their paper that monotonicity implies separability, where separability is meant to mean that an individual's weight may only depend on that individual's preferences, and not other agents' preferences.

assume that individual and social preferences satisfied the von Neumann-Morgenstern postulates. However, Harsanyi obtained a weaker conclusion than Fleming, namely only a single profile theorem, whereas Fleming's theorem was a multi-profile theorem.[11] Interestingly, the axiom that Harsanyi dropped[12] was a separability axiom that is somewhat similar to the separability axiom that we use, although it is phrased in terms of vectors of individual utilities, not in terms of ordinal preferences. One might thus view our work as integrating Fleming's and Harsanyi's approaches.

Related Literature

The invariance axiom in this paper is related to, but much weaker, than Arrow's [1] "independence of irrelevant alternatives." The main difference between these axioms concerns *what* is regarded as an irrelevant alternative. According to Arrow's axiom, for the comparison of any two alternatives, *all* other alternatives are irrelevant. By contrast, according to the invariance axiom, for the social preferences over all alternatives, only those alternatives are regarded as irrelevant for which all agents agree that they are equivalent to lotteries over the given subset.

If, instead of the invariance axiom, Arrow's independence of irrelevant alternatives axiom were used, one would obtain versions of Arrow's impossibility theorem. This is not obvious because we are considering a social welfare function with a smaller domain than Arrow's [1] social welfare function. Arrow considers a "full" domain, whereas we only consider expected utility preferences. But it was shown by Sen [26, Theorem 8*2] that Arrow's theorem remains valid on this restricted domain. Stronger versions of Sen's result were shown by Kalai and Schmeidler [21] and Hylland [20]. This literature sometimes refers to "cardinal utilities" rather than Bernoulli utilities, but the results that we have quoted, even if they refer to cardinal utilities, can also be interpreted as results about Bernoulli utilities. Note that these results imply that relative utilitarianism does not satisfy independence of irrelevant alternatives.

Chichilnisky [6] proved another impossibility result in this area, namely the non-existence of a continuous aggregation rule for von Neumann-Morgenstern preferences that also respects unanimity and that is anonymous. No-

[11] In a comment on Harsanyi's paper, Fleming [16] does not raise the issue of single-profile vs. multi-profile results.

[12] Although Harsanyi argued for the plausibility of the axiom [18, pp. 310-312].

tice that relative utilitarianism is not continuous. For example, if agent 1's Bernoulli utility over three alternatives is given by the vector $(1, 0.5 + \varepsilon, 0)$, and agent 2's Bernoulli utility for the same three alternatives is $(0, 0.5(1 - \varepsilon), 1)$, then for every $\varepsilon > 0$ the sum of these two utility vectors is $(1, 1 + 0.5\varepsilon, 1)$, which corresponds to the same social preferences as the vector $(1, 2, 1)$, but in the limit, as $\varepsilon \to 0$, the sum of the utility vectors corresponds to complete indifference. In our development here, we shall not impose any explicit continuity requirements, although our anonymity axiom implies a weak continuity requirement.

Dhillon and Mertens [10] proved another impossibility result in this area. They showed that the Pareto axiom and a strong form of monotonicity cannot be satisfied simultaneously. Note that this result implies that relative utilitarianism does not satisfy this strong form of monotonicity.

An axiomatization of relative utilitarianism that is very different from the one pursued in this paper was provided by Segal [25]. Whereas our paper follows Arrow's [1] approach and considers for a variety of lists of individuals' preferences how welfare is defined, holding the set of alternatives constant, Segal considers for a variety of sets of alternatives how welfare is defined, holding the individuals' preferences fixed.

Outline of the Paper

In Sections 2 and 3 we review Harsanyi's [18] aggregation theorem, and explain how a social preference that satisfies a Pareto Axiom reveals the marginal rates of substitution between the probabilities of different agents' most preferred alternatives, assuming that any increase/decrease in the probability of some agent's most preferred alternative is accompanied by an equally sized decrease/increase in the agent's least preferred alternative. In Section 4 we extend the framework and consider multi-profile social welfare functions. In Sections 5, 6, and 7 we successively introduce the three axioms which characterize relative utilitarianism, and discuss each axiom's implications. As we develop our argument, we shall make three assumptions regarding the domain of the social welfare function. These assumptions greatly simplify our arguments. In Section 8 we provide a result that asserts the existence of a domain that satisfies our assumptions, and that is dense in the set of all preference profiles that satisfy von Neumann-Morgenstern axioms. This result is proved in an appendix. Section 9 concludes.

2 The Pareto Axiom

There are a finite set of alternatives, $A = \{a_1, a_2, \ldots, a_m\}$, and a finite set of individuals $N = \{1, 2, \ldots, n\}$. We assume that both m and n are at least 2. We denote the set of all lotteries over A by ΔA. The set of all preference orderings over ΔA that satisfy the von Neumann-Morgenstern axioms will be denoted by $\overline{\mathcal{R}}$. Every individual $i \in N$ will be assumed to have a preference ordering $\succsim_i \in \overline{\mathcal{R}}$. We assume that no individual is indifferent between all lotteries. The set of all preference orderings over ΔA that satisfy the von Neumann-Morgenstern axioms and that are not indifferent between all lotteries will be denoted by $\mathcal{R}$. Thus, $\succsim_i \in \mathcal{R}$ for all $i \in N$. The assumption that no individual is indifferent between all alternatives is very mild. Individuals who are indifferent between all alternatives could arguably be dropped from the analysis. The strict preference derived from $\succsim_i$ will be denoted by $\succ_i$, and the indifference relation derived from $\succsim_i$ will be denoted by $\sim_i$.

In this and the next section we take as given and fixed a profile $\succsim = (\succsim_i)_{i \in N} \in \mathcal{R}^n$ of preferences, one for each individual. We seek to investigate a benevolent social planner's preference. We denote this preference by $\succsim_s$. We shall also refer to $\succsim_s$ as the "social preference." We assume that $\succsim_s$ satisfies the von Neumann-Morgenstern axioms. We allow for the possibility that $\succsim_s$ is indifferent between all alternatives. Thus, $\succsim_s \in \overline{\mathcal{R}}$. We denote by $\succ_s$ the strict preference order derived from $\succsim_s$ and by $\sim_s$ the indifference relation.

Definition 1. The social preference $\succsim_s$ satisfies the *Pareto Axiom with respect to* $\succsim$ if for all $p, q \in \Delta A$:

(i) If $p \succsim_i q$ for all $i \in N$, then $p \succsim_s q$.

(ii) If $p \succsim_i q$ for all $i \in N$, and $p \succ_i q$ for at least one $i \in N$, then $p \succ_s q$.

The following proposition, which is closely related to Harsanyi's [18] theorem on utilitarianism, is the first part of Theorem 3 in Weymark [28].

Theorem 1. *The following two conditions are equivalent:*
(i) $\succsim_s$ *satisfies the Pareto axiom with respect to* $\succsim$.
(ii) Whenever for every $i \in N$ $u_i : A \to \mathbb{R}$ *is a Bernoulli utility function that represents* $\succsim_i$, *and* $u_s : A \to \mathbb{R}$ *is a Bernoulli utility function that*

represents $\succsim_s$, then there are strictly positive real numbers w_i for all $i \in N$, and a real number μ, such that:

$$u_s(a) = \sum_{i \in N} w_i u_i(a) + \mu \text{ for all } a \in A.$$

3 Revealed Marginal Rates of Substitution

We now investigate for any two agents i and j whether the social preference relation $\succsim_s$ reveals how much "relative weight" the social preference attaches to agent i's and agent j's preferences. As our approach to relative utilitarianism in this paper is purely based on preferences, and not on their numerical representations, we shall define this "relative weight" in terms of the preferences only. We shall introduce a concept called "the social preference's marginal rate of substitution between agents i and j." This marginal rate of substitution indicates how much probability of agent j's most preferred alternative we can subtract and keep welfare constant if we raise the probability of agent i's most preferred alternative by one unit. Here, we shall assume that all subtractions (additions) from (to) the probability of an agent's most preferred alternative are accompanied by equal additions (subtractions) to (from) the probability of that agent's least preferred alternative, and we shall assume that all agents other than i and j are indifferent towards these changes in probability. If this marginal rate of substitution is large, then intuitively agent i's "relative weight" in comparison to agent j is large, whereas if the marginal rate of substitution is small, then intuitively agent i's "relative weight" in comparison to agent j is low.

To define marginal rates of substitution formally, we introduce some more notation. For any agent $i \in N$, we denote by b_i one of agent i's most preferred alternatives in A and by ℓ_i one of agent i's least preferred alternatives in A (in each case it does not matter which one we pick, if there are multiple most or least preferred alternatives). For any two alternatives $a, b \in A$ and for any $\lambda \in [0, 1]$ we write $\lambda a + (1 - \lambda)b$ for the lottery in ΔA that places probability λ on a and probability $1 - \lambda$ on b. Note that for any agent $i \in N$ and any lottery $q \in \Delta A$, because $\succsim_i$ is not indifferent between all elements of ΔA, there is a unique number $\alpha_i(q)$ such that $q \sim_i \alpha_i(q)b_i + (1 - \alpha_i(q))\ell_i$. We can now define marginal rates of substitution between agents i and j revealed by a social preference.

Definition 2. Suppose $\succsim \in \mathcal{R}^N$, $i, j \in N$, and $i \neq j$. Let $\succsim_s$ be the social preference. For any two lotteries $p, q \in \Delta A$ such that:

$$p \sim_s q$$

and

$$\begin{aligned} &p \sim_k q \quad \text{for all } k \in N \setminus \{i, j\} \\ &p \succ_i q, \\ &q \succ_j p, \end{aligned}$$

the social preference $\succsim_s$ *reveals that the marginal rate of substitution between i and j at p and q* is:

$$\text{MRS}_{i,j}(p, q) = \frac{\alpha_i(p) - \alpha_i(q)}{\alpha_j(q) - \alpha_j(p)}$$

This definition of the revealed marginal rate of substitution is based on a movement from some lottery p to another lottery q, and, for the moment, the value of the marginal rate of substitution may depend on which lotteries p and q we are considering. As in any definition of marginal rates of substitution we consider movements along a indifference curve; this is expressed in the definition by the assumption $p \sim_s q$. Because we want to focus on the marginal rate of substitution between agents i and j, we require that all other agents are indifferent between p and q. Finally, as we are interested in how agent i's and agent j's preferences are traded off against each other, we assume that i and j have strict and opposite preferences over p and q. The marginal rate of substitution between i and j at p and q is then defined as the change in the probability of the most preferred alternative of i that is for i equivalent to the movement from q to p divided by the same change, reversing the order of p and q, for j. Thus, the marginal rate of substitution indicates by how much probability of agent j's most preferred alternative we can subtract and keep welfare constant if we raise the probability of agent i's most preferred alternative by one unit. Here it is assumed that any increase/decrease in the probability of an agent's preferred alternative is matched by a equally sized decrease/increase in the probability of that agent's least preferred alternative.

Before we can use the concept of revealed marginal rate of substitution, we have to address whether such rates always exist, i.e. whether we can find at least one pair of suitable lotteries p and q, and whether, if several such pairs of lotteries exist, the marginal rate of substitution is independent of which pair we pick, that is, whether it is unique. We begin with existence. Not every social preference that satisfies the Pareto axiom reveals a marginal rate of substitution. Suppose, for example, that two individuals have identical preferences. Then, regardless of the other agents' preferences, and regardless of what the social preference is, it will be impossible to reveal a marginal rate of substitution that involves either of these two individuals because lotteries satisfying the conditions of Definition 2 don't exist.

In addition to the preference profiles in the previous paragraph, there are also profiles $\succsim$ such that *some, but not all* social preferences that satisfy the Pareto axiom reveal a marginal rate of substitution between i and j. Here is an example. Suppose society consists of just two individuals, 1 and 2, and there are just two alternatives, a and b. Suppose the preference profile is such that 1 prefers a to b but 2 prefers b to a. The social preference where society is indifferent between a and b reveals that the marginal rate of substitution between 1 and 2 is 1. In contrast, the social preference where society prefers a to b fails to reveal a marginal rate of substitution between 1 and 2, because there do not exist lotteries p and q such that 1 prefers the former, 2 prefers the latter, and society is indifferent between the two.

We now ask: For which preference profiles $\succsim$ does *any* social preference $\succsim_s$ that satisfies the Pareto axiom reveal at least one marginal rate of substitution between i and j for all $i, j \in N$ with $i \neq j$? This is a relevant question in our context because for such profiles the axioms that we introduce below do not have to be conditioned on the social preference revealing a marginal rate of substitution. We are assured that at least one marginal rate of substitution is revealed. A sufficient condition for this to be the case was introduced by Fishburn [14], who used it for a slightly different purpose than we do. Weymark [28] introduced the name "Independent Prospects" for this condition. One may interpret this condition as saying that for every individual i in society there is at least one pair of lotteries such that the difference between these lotteries is a private matter of that individual, and is of no concern to any other individual.

Definition 3. A profile of preferences $\succsim \in \mathcal{R}^n$ satisfies the *Independent Prospects* condition if for every $i \in N$ there are lotteries $p_i, q_i \in \Delta A$ such that $p_i \succ_i q_i$, and $p_i \sim_k q_i$ for all $k \in N \setminus \{i\}$.

Observe that the independent prospects condition cannot be satisfied if two agents have the same preferences. Note also that the independent prospects condition can only be satisfied if there are more alternatives than individuals: $m > n$. We will later restrict attention to profiles that satisfy the independent prospects condition, and therefore, we will assume in particular that no individuals ever have exactly identical preferences, and that $m > n$.

Proposition 1. *Suppose $\succsim$ satisfies the Independent Prospects condition, and suppose that the social preference $\succsim_s$ satisfies the Pareto axiom. Then $\succsim_s$ reveals a marginal rate of substitution between every pair $i, j \in N, i \neq j$ of agents.*

Proof. Let $i, j \in N, i \neq j$. To prove the Proposition, it suffices to construct lotteries p, q that satisfy the conditions in Definition 2. We start with the lotteries p_i, q_i, p_j, q_j whose existence is given by Definition 3. For any $\alpha \in [0, 1/4]$ let:

$$p(\alpha) = \left(\frac{1}{2} - \alpha\right) p_i + \alpha q_i + \left(\frac{1}{4} - \alpha\right) p_j + \left(\frac{1}{4} + \alpha\right) q_j. \tag{1}$$

$$q(\alpha) = \alpha p_i + \left(\frac{1}{2} - \alpha\right) q_i + \left(\frac{1}{4} + \alpha\right) p_j + \left(\frac{1}{4} - \alpha\right) q_j. \tag{2}$$

If $\alpha = 0$, then $p(\alpha) \succ_i q(\alpha)$ and $p(\alpha) \sim_k q(\alpha)$ for all $k \neq i$, so that by the Pareto axiom, $p(\alpha) \succ_s q(\alpha)$. Conversely, if $\alpha = 1/4$, then $q(\alpha) \succ_j p(\alpha)$ and $q(\alpha) \sim_k p(\alpha)$ for all $k \neq j$, so that by the Pareto axiom, $q(\alpha) \succ_s p(\alpha)$. Hence, by the continuity of von Neumann-Morgenstern preferences, there exists $\bar{\alpha} \in (0, 1/4)$ such that $p(\bar{\alpha}) \sim_s q(\bar{\alpha})$. We now set $p \equiv p(\bar{\alpha})$ and $q \equiv q(\bar{\alpha})$. By construction $p \sim_s q$. The claim is proved if we show that $p \succ_i q$ and $q \succ_j p$, which follows from the fact that $\bar{\alpha}$ is in the interior of $[0, 1/4]$, and that $p \sim_k q$ for all $k \neq i, j$, which is true because, for such agents k, we have $p(\alpha) \sim_k q(\alpha)$ for all $\alpha \in [0, 1/4]$. □

Suppose next that a social preference that satisfies the Pareto axiom does reveal at least one marginal rate of substitution between agents i and j. Is this marginal rate of substitution uniquely determined, or could several values of the marginal rate of substitution be revealed? It is one of the implications of the following result that the marginal rate of substitution is uniquely determined.

But first, some notation: For any $\succsim_i \in \mathcal{R}$ let $u(\succsim_i)$ denote the Bernoulli utility function that represents $\succsim_i$ and that is normalized: $u(\succsim_i)(b_i) = 1$ and $u(\succsim_i)(\ell_i) = 0$. If it is clear from the context that individual i's preference relation is $\succsim_i$, then we shall write $u_i(a)$ instead of $u(\succsim_i)(a)$.

Proposition 2. *Consider a given preference profile $\succsim \in \mathcal{R}^n$, and let $\succsim_s \in \overline{\mathcal{R}}$ be the corresponding social preference. Suppose that $\sum_{i \in N} w_i u(\succsim_i)$ is a Bernoulli utility function that represents $\succsim_s$, where each $w_i \in \mathbb{R}_{++}$. Suppose that the social preference $\succsim_s$ reveals that the marginal rate of substitution between i and j at some pair of lotteries p and q to be $MRS_{i,j}(p,q)$. Then:*

$$MRS_{i,j}(p,q) = \frac{w_j}{w_i}.$$

There may be multiple pairs of lotteries p, q that satisfy the conditions of Definition 2. But under the conditions of Proposition 2 for all such pairs of lotteries the revealed marginal rate of substitution equals w_j/w_i, and therefore the revealed marginal rate of substitution is unique.

Proof. Because the social preference is indifferent between p and q:

$$\sum_{k \in N} w_k u_k(p) = \sum_{k \in N} w_k u_k(q) \tag{3}$$

Because agents other than i and j are indifferent between p and q, this is equivalent to:

$$\sum_{k \in \{i,j\}} w_k u_k(p) = \sum_{k \in \{i,j\}} w_k u_k(q) \tag{4}$$

which simplifies to:

$$\frac{u_i(p) - u_i(q)}{u_j(q) - u_j(p)} = \frac{w_j}{w_i} \tag{5}$$

Because the utility functions u_k are normalized so that the utility of $u_k(b_k) = 1$ and $u_k(w_k) = 0$, we can replace $u_k(p)$ by $\alpha_k(p)$ and $u_k(q)$ by $\alpha_k(q)$ for $k = i, j$, and we obtain the desired result. □

4 Social Welfare Functions

We now consider preference aggregation not only for one preference profile, but for every preference profile in some set of preference profiles. The obvious choice of the set of preference profiles to be considered is $\mathcal{R}^n$, but in this paper we shall restrict attention to a subset $\mathscr{R}$ of $\mathcal{R}^n$. The reason is that we only want to consider preference profiles that satisfy the independent prospects condition, so that we can be sure that a social preference that satisfies the Pareto axiom reveals marginal rates of substitution. Our axioms later in the paper will be terms of the revealed marginal rates of substitution.

We don't specify the set $\mathscr{R}$ here, but rather throughout the paper we will make assumptions regarding this set as we use them, and in Section 8 and in the appendix we shall construct an example of a domain that satisfies all our assumptions. The set $\mathscr{R}$ will later be assumed to be topologically dense in $\mathcal{R}^n$. Our first assumption is:

Assumption 1. *$\mathscr{R}$ is the Cartesian product of non-empty sets of preferences for each agent. That is, $\mathscr{R} = \times_{i\in N}\mathscr{R}_i$, where for each $i \in N$, $\emptyset \neq \mathscr{R}_i \subset \mathcal{R}$. Moreover, every $\succsim \in \mathscr{R}$ satisfies the Independent Prospects condition.*

Notice the requirement in Assumption 1 that $\mathscr{R}$ is a Cartesian product. If it were not a Cartesian product, the set of preferences of some agent i that we consider would depend on the preferences of all other agents. In other words, our study of preference aggregation would implicitly assume a form of correlation among agents' preferences. We see no good intuitive reason to introduce such a correlation. Moreover, the Cartesian product assumption simplifies our terminology and notation and makes our main arguments, for example in the next section, easier to follow. On the other hand, this assumption complicates our construction of the set $\mathscr{R}$ in the appendix. For us the transparency of the arguments in the main text of the paper is more important.

Recall that the Independent Prospects condition cannot be satisfied if two agents have the same preferences. Therefore, Assumption 1 implies that for any two i, j with $i \neq j$ we have: $\mathscr{R}_i \cap \mathscr{R}_j = \emptyset$. Recall also that the Independent Prospects condition can only be satisfied if there are more alternatives than agents. Thus, our paper relies on the assumption: $m > n$.

Definition 4. A *social welfare function* (SWF) is a function:

$$\phi : \mathscr{R} \longrightarrow \overline{\mathcal{R}}.$$

Our focus will be on the SWFs that satisfy the *Pareto Axiom.* We extend this axiom from single profiles to SWFs as follows.

Definition 5. A SWF ϕ satisfies the *Pareto Axiom* if for all preference profiles $\succsim \in \mathscr{R}$, the social preference $\phi(\succsim)$ satisfies the Pareto Axiom with respect to $\succsim$.

Let ϕ be a SWF that satisfies the Pareto Axiom. Then for every $\succsim \in \mathscr{R}$ and all $i, j \in N$ with $i \neq j$, we can identify the revealed marginal rate of substitution for i and j. We denote these marginal rate of substitution by $\text{MRS}_{i,j}(\succsim)$. In the next three sections we shall consider the implications of three axioms regarding the marginal rates of substitution.

5 Separability of Revealed Marginal Rates of Substitution

Our axiomatization of relative utilitarianism will focus on the marginal rates of substitution revealed by a utilitarian welfare function. The first axiom is separability.

Definition 6. Suppose that the domain $\mathscr{R}$ of a SWF ϕ satisfies Assumption 1, and that ϕ itself satisfies the Pareto Axiom. Then we say that ϕ satisfies in addition also the *Separability Axiom* if for all $i, j \in N$ with $i \neq j$ and for all $\succsim, \hat{\succsim} \in \mathscr{R}$ such that $\succsim_i = \hat{\succsim}_i$ and $\succsim_j = \hat{\succsim}_j$ we have:

$$\text{MRS}_{i,j}(\succsim) = \text{MRS}_{i,j}(\hat{\succsim})$$

Separability is implied by Axiom 1 in Dhillon [9].[13] The idea underlying this axiom offers one possible motivation for requiring separability. Suppose, instead of aggregating the preferences of all agents in N simultaneously, we proceeded in two steps: First, we aggregated the preferences of the sub-group consisting of only two individuals, i and j, and then we treated the subgroup as if it was one individual with preference equal to the social preference of the subgroup, and aggregated this artificial individual's preference and the preferences of all individuals in $N \setminus \{i, j\}$. Dhillon's

[13]The idea of this axiom also appears in Dhillon and Mertens [11, p.481-2].

Axiom 1 requires this two step procedure to lead to the same social preference as the aggregation of all agents' preferences simultaneously.[14] She formalizes this by postulating that for any subset of N there is a social welfare function that assigns to each vector of preferences of the individuals in this subset a social preference, and by postulating that these social welfare functions are consistent with each other in the sense that for any group of agents the social preference could be obtained by partitioning this group into subsets, aggregating each subset's preferences separately according to the social welfare function for that subset, and then aggregating the preferences that one has obtained in that way. Moreover, she requires that each social welfare function satisfies the Pareto Axiom, and thus has the utilitarian form. Our Separability Axiom restricts attention to groups of two. It is an implication of Dhillon's axiom because implicit in Dhillon's construction is that the social welfare function for the group consisting of i and j is independent of the preferences of the other members of N.

As mentioned in the Introduction, separability was also a key axiom in Fleming's [15] axiomatization of utilitarianism. Harsanyi [18] provided an eloquent defense of separability, although he did not use it in his own theorem. Harsanyi draws a parallel with the Pareto axiom, and writes that "both postulates make social choice dependent solely on the *individual* interests directly affected. They leave no room for the separate interests of a superindividual state or of impersonal cultural values ..." [18, p. 311].

Theorem 2. *Suppose $n \geq 3$ and let the domain $\mathscr{R}$ of the social welfare function ϕ satisfy Assumption 1. Then ϕ satisfies the Pareto and the Separability Axioms if and only if for every $i \in N$ there are functions $\lambda_i : \mathscr{R}_i \to \mathbb{R}_{++}$ such that for every $\succsim \in \mathscr{R}$ the social preference $\phi(\succsim)$ can be represented by:*

$$u_s = \sum_{i \in N} \lambda_i(\succsim_i) u(\succsim_i).$$

Proof. The "if part" is obvious. We prove the "only if part". For every $\succsim \in \mathscr{R}$ let the Bernoulli utility function $\sum_{i \in N} w_i(\succsim) u(\succsim_i)$ represent $\phi(\succsim)$. In the following proof we shall construct the weights $\lambda_i(\succsim_i)$ the existence of which is asserted in Theorem 2. To be able to appeal to standard results

[14] Axiom 1 in Dhillon [9] is not phrased as we describe it here. But Dhillon's comments following Axiom 1 indicate that her Axiom 1 is equivalent to the condition that we describe.

on additive separability we shall use the logarithms of the weights in the welfare function. We define for every $\succsim \in \mathscr{R}$ and every $i \in N$:

$$v_i(\succsim) = \ln w_i(\succsim). \tag{6}$$

By Proposition 2: $MRS_{i,j}(\succsim) = v_i(\succsim) - v_j(\succsim)$. The Separability Axiom implies for all $i, j \in N$:

$$v_i(\succsim) - v_j(\succsim) = v_i(\hat{\succsim}) - v_j(\hat{\succsim}) \tag{7}$$

whenever $\succsim_i = \hat{\succsim}_i$ and $\succsim_j = \hat{\succsim}_j$. Define for every $i \in N$ a function $h_i : \mathscr{R}_i \times \mathscr{R}_{i+1} \to \mathbb{R}$ such that:

$$h_i(\succsim_i, \succsim_{i+1}) = v_{i+1}(\succsim) - v_i(\succsim) \tag{8}$$

where, because of the Separability Axiom, it does not matter which preference profile $\succsim$ we consider as long as i's preference in this profile is $\succsim_i$, and $i+1$'s preference in the profile is $\succsim_{i+1}$. We can extend this definition to the case $i = n$ by identifying $n+1$ with 1.

Now notice that for all $\succsim \in \mathscr{R}$ we have:

$$v_1(\succsim) = v_1(\succsim) + \sum_{i=1}^{n} h_i(\succsim_i, \succsim_{i+1}) \Leftrightarrow \tag{9}$$

$$\sum_{i=1}^{n} h_i(\succsim_i, \succsim_{i+1}) = 0. \tag{10}$$

This implies that for any $\succsim, \hat{\succsim} \in \mathscr{R}$ we have:

$$\sum_{i=1}^{n} h_i(\succsim_i, \succsim_{i+1}) = \sum_{i=1}^{n} h_i(\hat{\succsim}_i, \hat{\succsim}_{i+1}). \tag{11}$$

In the special case in which $\succsim_j = \hat{\succsim}_j$ for all j except one i, this equation can be simplified by dropping all terms that appear on both sides of the equation. We then obtain:

$$h_{i-1}(\succsim_{i-1}, \succsim_i) - h_{i-1}(\succsim_{i-1}, \hat{\succsim}_i) = h_i(\hat{\succsim}_i, \succsim_{i+1}) - h_i(\succsim_i, \succsim_{i+1}) \tag{12}$$

Because $n \geq 3$, we know that $i-1 \neq i+1$. This means, that in this equation the left hand side must not depend on $\succsim_{i-1}$, because this preference does not appear on the right hand side. This applies in fact to all i and all $\succsim \in \mathscr{R}$. Thus the increments of the function h_i when the second argument is changed, must not depend on the first argument, and also the increments of the function h_i, when the first argument is changed, must not depend on the second argument. These conditions imply by standard arguments that the functions h_i are additively separable, i.e. there exist functions $f_i : \mathscr{R}_i \to \mathbb{R}$ and $g_i : \mathscr{R}_i \to \mathbb{R}$ such that:

$$h_i(\succsim_i, \succsim_{i+1}) = f_i(\succsim_i) + g_{i+1}(\succsim_{i+1}) \tag{13}$$

for all $i \in N$ and all $\succsim \in \mathscr{R}$.

Plugging equation (13) into equation (10) we get:

$$\sum_{i=1}^{n} (f_i(\succsim_i) + g_{i+1}(\succsim_{i+1})) = 0, \tag{14}$$

which is, of course, the same equation as:

$$\sum_{i=1}^{n} (f_i(\succsim_i) + g_i(\succsim_i)) = 0. \tag{15}$$

This equation can be true for all $\succsim \in \mathscr{R}$ only if each of the terms in the sum on the left hand side is a constant that is independent of $\succsim_i$, i.e. there is some $k_i \in \mathbb{R}$ such that:

$$f_i(\succsim_i) + g_i(\succsim_i) = k_i \tag{16}$$

for every $\succsim_i \in \mathscr{R}_i$. Using this, we can re-write (13) as:

$$h_i(\succsim_i, \succsim_{i+1}) = f_i(\succsim_i) + k_{i+1} - f_{i+1}(\succsim_{i+1}). \tag{17}$$

Substituting this into (8) we obtain:

$$v_{i+1}(\succsim) - v_i(\succsim) = f_i(\succsim_i) + k_{i+1} - f_{i+1}(\succsim_{i+1}). \tag{18}$$

Now we return to the original variables that we are interested in, rather than their logarithms. We define for every $i \in N$ and for every $\succsim_i \in \mathscr{R}_i$:

$$\psi_i(\succsim_i) = \exp(-f_i(\succsim_i)), \tag{19}$$

and:

$$\alpha_i = \exp(k_i). \tag{20}$$

We can now apply the exponential function to both sides of (18) and get:

$$\frac{w_{i+1}(\succsim)}{w_i(\succsim)} = \alpha_{i+1}\frac{\psi_{i+1}(\succsim_{i+1})}{\psi_i(\succsim_i)}. \tag{21}$$

Now if we define for every $i \in N$:

$$\lambda_i(\succsim_i) = \alpha_1 \cdot \alpha_2 \cdot \ldots \cdot \alpha_i \cdot \psi_i(\succsim_i), \tag{22}$$

then:

$$\frac{\lambda_{i+1}(\succsim_{i+1})}{\lambda_i(\succsim_i)} = \alpha_{i+1}\frac{\psi_{i+1}(\succsim_{i+1})}{\psi_i(\succsim_i)}, \tag{23}$$

and thus the vector $(\lambda_1, \ldots, \lambda_n)$ is proportional to the vector $(w_1, \ldots, w_n)$. Therefore,

$$u_s = \sum_{i \in N} \lambda_i(\succsim_i) u(\succsim_i) \tag{24}$$

is a representation of the social preference. □

6 Invariance of Marginal Rates of Substitution

We now introduce our third axiom. This axiom, together with the previous two axioms, implies that the marginal rates of substitution remain the same across all preference profiles. In this axiom, if $A' \subseteq A$, we denote by $\succsim_i |_{\Delta(A')}$ the restriction of the preference relation $\succsim_i$ to lotteries that have support in A'.

Definition 7. A SWF ϕ that satisfies the Pareto Axiom, satisfies in addition also the *Invariance Axiom* if for all $i \in N$, $\succsim, \hat{\succsim} \in \mathscr{R}$, and $a, b, c \in A$ (all different from each other) the following holds. If

- $\succsim_j = \hat{\succsim}_j$ for all $j \in N \setminus \{i\}$,
- $a \sim_j b \sim_j c$ for all $j \in N \setminus \{i\}$,
- $\succsim_i|_{\Delta(A\setminus\{b\})} = \hat{\succsim}_i |_{\Delta(A\setminus\{b\})}$,

- $a \succsim_i b \succsim_i c$ and $a \,\hat{\succsim}_i\, b \,\hat{\succsim}_i\, c$,

then:

$$\mathrm{MRS}_{i,j}(\succsim) = \mathrm{MRS}_{i,j}(\hat{\succsim}) \text{ for all } j \in N, j \neq i.$$

In words, the axiom requires that, under certain conditions, the marginal rates of substitution involving agent i don't change if agent i's preference alone changes while all other agents' preferences stay the same. If we required this regardless of what agent i's and all other agents' preferences are, then we would assume our intended conclusion, as long as we also imposed the Pareto and the Separability axioms. However, it is sufficient to require invariance of the marginal rates of substitution under much more restrictive conditions, namely those listed in the bullet points in Definition 7. These conditions are that there are alternatives a, b and c such that (i) all agents other than i are indifferent between a, b and c, whereas agent i ranks b between a and c, and (ii) only i's preferences regarding lotteries assigning positive probability to b change, leaving his preferences on $\Delta(A \setminus \{b\})$ unchanged, and also leaving unchanged that b is ranked between a and c.

Why is it interesting to explore the implications for the social welfare function of the assumption that, in these circumstances, marginal rates of substitution don't change? Before the change of preference, there is a lottery over a and c such that all agents are indifferent between that lottery and the alternative b. After the change of preference, there is some (potentially different) lottery over a and c such that the same is true. Hence, both before and after the change of agent i's preference, alternative b is, in the words of Dhillon and Mertens, "redundant." It might be plausible to argue that i's preferences over a redundant alternative should not affect the marginal rates of substitution involving i.

Dhillon and Mertens motivate their version of the invariance axiom by pointing out that the preference change that is considered in the invariance axiom leaves the set of vectors of expected utilities that correspond to lotteries over A unchanged. Thus, if one assumes that what matters for welfare is only the image of the choice space in expected utility space, then the Invariance Axiom follows.

The Invariance Axiom has bite only if the domain of the welfare function is sufficiently rich. Assumption 2 below ensures this richness. We first need a definition:

	$\succsim_i$	$\succsim_i^2$	$\succsim_i^3$	$\succsim_i^4$	$\succsim_i^5$	$\hat{\succsim}_i$
a	1	1	0.1	0.1	0.1	0.1
b	0.6	0.6	0.6	0	0	0
c	0.4	1	1	1	1	1
d	0.2	0.2	0.2	0.2	0.2	0.5
e	0	0	0	0	0.2	0.2

Figure 1: $\hat{\succsim}_i \in \mathscr{R}_i$ can be reached from $\succsim_i \in \mathscr{R}_i$ through a sequence of simple modifications.

Definition 8. A "simple modification" of a preference $\succsim_i \in \mathscr{R}_i$ is a preference $\hat{\succsim}_i \in \mathscr{R}_i$ such that $u(\hat{\succsim}_i)$ assigns the same utility to all alternatives in A as $u(\succsim_i)$ except to one alternative $b \in A$, and moreover such that there is an alternative $a \neq b$ to which $u(\succsim_i)$ assigns 1, and also an alternative $c \neq b$ to which $u(\succsim_i)$ assigns 0. We say that "a preference relation $\hat{\succsim}_i \in \mathscr{R}_i$ can be reached from a preference relation $\succsim_i \in \mathscr{R}_i$ through a sequence of simple modifications" if there is a sequence $(\succsim_i^k)_{k=1,2,\ldots,K}$ of elements of $\mathscr{R}_i$ such that $\succsim_i^1 = \succsim_i$, $\succsim_i^K = \hat{\succsim}_i$, and for every $k = 1, 2, \ldots, K-1$ the preference relation $\succsim_i^{k+1}$ is a simple modification of the preference relation $\succsim_i^k$.

In Figure 1 we illustrate how one preference can be reached from another through simple modifications. Each row corresponds to an alternative, each column corresponds to a preference, and preferences are represented by von Neumann-Morgenstern utility functions the values of which constitute the entries in the table in Figure 1. The sequence of simple modifications by which $\hat{\succsim}_i$ is reached from $\succsim_i$ proceeds from the left to the right.

The starting and end points of the sequence in Figure 1 have been chosen quite arbitrarily. This is to suggest that it is in fact easy to connect any pair of preferences through a sequence of simple modifications as long as the sets $\mathscr{R}_i$ are sufficiently large. Implicitly, part (i) of Assumption 2 below is therefore a richness assumption for the domain of the SWF that we are considering.

Assumption 2. *(i) For every $i \in N$, every preference $\hat{\succsim}_i \in \mathscr{R}_i$ can be* *reached from every other preference $\succsim_i \in \mathscr{R}_i$ through a sequence of simple modifications.*

(ii) For every $i \in N$, for any three alternatives $a, b, c \in A$, there is a preference $\succsim_i \in \mathscr{R}_i$ such that $a \sim_i b \sim_i c$.

Obviously, condition (ii), and our assumption that agents are not indifferent between all alternatives, imply that there must be at least four alternatives. The following theorem assumes $n \geq 3$, and also Assumption 1, which, as mentioned earlier, implies $m > n$. Thus, these assumptions alone imply that $m \geq 4$. Thus, condition (ii) in Assumption 2 does not introduce additional restrictions regarding the number of alternatives.

Theorem 3. *Suppose $n \geq 3$ and suppose that $\mathscr{R}$ satisfies Assumptions 1 and 2. Then a SWF ϕ satisfies the Pareto, Separability, and Invariance Axioms if and only if for every $i \in N$ there is a number $\lambda_i \in \mathbb{R}_{++}$ such that for every $\succsim \in \mathscr{R}$ the social preference $\phi(\succsim)$ can be represented by:*

$$u_s = \sum_{i \in N} \lambda_i u(\succsim_i).$$

Proof. The "if part" is obvious. We prove the "only if part". By Theorem 2 there are functions $\lambda_i : \mathscr{R}_i \to \mathbb{R}_{++}$ such that for every $\succsim \in \mathscr{R}$ the social preference $\phi(\succsim)$ can be represented by: $u_s = \sum_{i \in N} \lambda_i(\succsim_i) u(\succsim_i)$. It remains to show that for every $i \in N$ and all $\succsim_i, \hat{\succsim}_i$ we have: $\lambda_i(\succsim_i) = \lambda_i(\hat{\succsim}_i)$. By part (i) of Assumption 2 $\hat{\succsim}_i$ can be reached from $\succsim_i$ through a sequence of simple modifications $(\succsim_i^k)_{k=1,2,\ldots,K}$. We shall prove the claim by showing that for every $k = 1, 2, \ldots, K-1$ we have $\lambda_i(\succsim_i^k) = \lambda_i(\succsim_i^{k+1})$.

We first construct for given $k \in \{1, 2, \ldots, K-1\}$ a preference profile $(\succsim_j^k)_{j \neq i} \equiv \succsim_{-i}^k$ such that the Invariance Axiom applies when agent i's preference changes from $\succsim_i^k$ to $\succsim_i^{k+1}$ while all other agents' preferences remain $\succsim_{-i}^k$. Denote by b the alternative whose utility changes when agent i's preferences switch from $\succsim_i^k$ to $\succsim_i^{k+1}$, denote by a an alternative other than b that is ranked top by $\succsim_i^k$, and by c an alternative other than b that is ranked bottom by $\succsim_i^k$. These alternatives exist because $\succsim_i^{k+1}$ is a simple modification of $\succsim_i^k$. For every $j \neq i$ we now pick some preference $\succsim_j^k \in \mathcal{R}_j$ such that $a \sim_j^k b \sim_j^k c$. Part (ii) of Assumption 2 implies that such a preference exists. Let $\succsim_{-i}^k$ be the list of the preferences $\succsim_j^k$ for all $j \neq i$.

The Invariance Axiom implies that for every $j \neq i$ the marginal rate of substitution for agent i and agent j is the same for $(\succsim_i^k, \succsim_{-i}^k)$ and $(\succsim_i^{k+1}, \succsim_{-i}^k)$. This is the case if and only if $\lambda_i(\succsim_i^k) = \lambda_i(\succsim_i^{k+1})$. □

7 Anonymity

We now add an anonymity axiom to obtain the conclusion that all marginal rates of substitution must equal 1, and therefore that the SWF must be Relative Utilitarian. A natural definition of anonymity may seem to be the requirement that for all preference profiles $(\succsim_k)_{k \in N}$, all permutations π of N, and all $i, j \in N, i \neq j$ the marginal rate of substitution $MRS_{i,j}((\succsim_k)_{k \in N})$ equals the marginal rate of substitution $MRS_{\pi(i),\pi(j)}((\succsim_{\pi(k)})_{k \in N})$. Unfortunately, given our domain restrictions, any two individuals' sets of possible preferences are disjoint, so that for any non-trivial permutation π if $(\succsim_k)_{k \in N} \in \mathscr{R}$ then $(\succsim_{\pi(k)})_{k \in N} \notin \mathscr{R}$.

We shall instead work with an "approximate" version of anonymity. Roughly speaking, it will require that if $(\succsim_k)_{k \in N} \in \mathscr{R}$, if π is a permutation of N, and if $(\hat{\succsim}_{\pi(k)})_{k \in N} \in \mathscr{R}$ is "close to" $(\succsim_{\pi(k)})_{k \in N}$, then the marginal rate of substitution $MRS_{i,j}((\succsim_k)_{k \in N})$ is "close to" the marginal rate of substitution $MRS_{\pi(i),\pi(j)}((\hat{\succsim}_{\pi(k)})_{k \in N})$ for all $i, j \in N, i \neq j$.

To formalize this requirement, we need to introduce a metric on von Neumann-Morgenstern preferences. We shall define the distance between two preferences $\succsim, \hat{\succsim} \in \mathcal{R}$ to be the Euclidean distance between their normalized von Neumann-Morgenstern representations (which we interpret here as vectors in $\mathbb{R}^m$): $||u(\succsim) - u(\hat{\succsim})||$. For our definition of anonymity it only matters which sequences of preferences are convergent. It is simple to verify that a sequence $(\succsim^\nu)_{\nu \in \mathbb{N}}$ of elements of $\mathcal{R}$ converges to $\succsim \in \mathcal{R}$ if and only if the upper contour sets of $\succsim^\nu$ converge in Hausdorff distance to the upper contour sets of $\succsim$. Thus, the notion of convergence that we use can be defined in purely ordinal terms.

Definition 9. A SWF ϕ that satisfies the Pareto axiom satisfies in addition the *Anonymity* axiom if for any preference profile $\succsim \in \mathscr{R}$, any permutation π of N, and any sequence of preference profiles $(\succsim^\nu)_{\nu \in \mathbb{N}}$ in $\mathscr{R}$ such that $\succsim_{\pi(i)}^\nu \rightarrow \succsim_i$ for all $i \in N$ we have:

$$MRS_{\pi(i),\pi(j)}(\succsim^\nu) \rightarrow MRS_{i,j}(\succsim) \tag{25}$$

for all $i, j \in N$, $i \neq j$.

Whether our formalization of anonymity has bite depends on the richness of the domain of the SWF. If the domain is finite, for example, then anonymity, as defined above, will always be satisfied, because no sequence of preferences of some agent i will ever converge to a preference of some other agent j. Therefore, to derive any further implications from the additional condition of anonymity, we have to make a richness assumption for the domain. We shall make a very strong assumption that allows a simple argument.

Assumption 3. *For every $i \in N$ the set of possible preferences of agent i, $\mathscr{R}_i$, is a dense subset of the set $\mathcal{R}$ of all von Neumann-Morgenstern preferences over lotteries over A.*

Theorem 4. *Suppose $n \geq 3$, and that $\mathscr{R}$ satisfies Assumptions 1, 2 and 3. Then a SWF ϕ satisfies the Pareto, Separability, Invariance, and Anonymity axioms if and only if for every $\succsim \in \mathscr{R}$ the social preference $\phi(\succsim)$ can be represented by:*

$$u_s = \sum_{i \in N} u(\succsim_i).$$

Proof. The "if part" of the result is obvious. We only prove the "only if part." Assumption 3 implies that for any preference profile $\succsim \in \mathscr{R}$ and any permutation π of N, there is a sequence of preference profiles $(\succsim^\nu)_{\nu \in \mathbb{N}}$ in $\mathscr{R}$ such that $\succsim^\nu_{\pi(i)} \rightarrow \succsim_i$ for all $i \in N$. From Theorem 3 and Proposition 2 we can infer that the sequence of marginal rates of substitution on the left hand side of condition (25) converges to $\lambda_{\pi(i)}/\lambda_{\pi(j)}$ whereas the marginal rate of substitution on the right hand side of that condition equals λ_i/λ_j. We conclude that to satisfy the anonymity axiom we must have for every permutation π of N that $\lambda_{\pi(i)}/\lambda_{\pi(j)} = \lambda_i/\lambda_j$ for all $i, j \in N, i \neq j$. But this implies $\lambda_i = \lambda_j$ for all $i, j \in N$, and hence, without loss of generality, we can set $\lambda_i = 1$ for all $i \in N$. This implies that the social preference can be represented by the utility function shown in the theorem. □

Our formalization of anonymity in this section appears to be closely related to the requirement that the SWF be continuous. This raises the question how our result is compatible with the impossibility result due to Chichilnisky [6] that we cited in the Introduction, and in which the continuity axiom is crucial to the result. The key point is that continuity in

Chichilnisky's work refers to the way in which the social preference itself depends on the individuals' preference profile. By contrast, in our definitions above we refer to the way in which the marginal rates of substitution revealed by the social preference depend on the individuals' preference profile.

8 Constructing the Domain

In this section we demonstrate the existence of an example of a domain $\mathscr{R}$ of a social welfare function that satisfies all assumptions that we have made in this paper. We go, in fact, one step further and also show that one can construct such a domain that is dense in the full domain $\mathcal{R}^n$. Denseness is interesting because it emphasizes that our construction does not leave any "holes" in the set of all preference profiles. To make this claim precise, we have to endow $\mathcal{R}^n$ with a metric. The metric that we use is defined by setting the difference between two preferences $\succsim_i$ and $\hat{\succsim}_i$ equal to the Euclidean distance of their normalized utility representations: $||u(\succsim_i) - u(\hat{\succsim}_i)||$. The metric on $\mathcal{R}^n$ is then the product metric.

The following result will be proved in the appendix, where we construct explicitly a domain that has all the properties listed in the Proposition. The construction is simple, but assumes that the number of alternatives is "much larger" than the number of agents: $m > 6n$. Recall that so far, our assumptions regarding the number of alternatives and agents have been: $m > n \geq 3$.

Proposition 3. *If $m > 6n$, then there exists at least one set $\mathscr{R} \subseteq \mathcal{R}^n$ that is the union of sets of preference profiles that each satisfy Assumptions 1, 2, and 3.*

One might ask whether Proposition 3 would remain true if we also required $\mathscr{R}$ to be open. Unfortunately, if $\mathscr{R} = \times_{i \in N} \mathscr{R}_i$ is open and dense in $\mathcal{R}^n$, then there is a preference profile $(\succsim_i)_{i \in N} \in \mathscr{R}$ which violates Independent Prospects. To see this note that if $\mathscr{R}$ is open and dense in $\mathcal{R}^n$ then also each $\mathscr{R}_i$ is open and dense in $\mathcal{R}$. Pick any $\succsim_1 \in \mathcal{R}_1$. Since $\mathcal{R}_1$ is open, there exists an open ball B around $\succsim_1$ with $B \subseteq \mathscr{R}_1$. Now, since $\mathscr{R}_2$ is dense in $\mathcal{R}$, it must be that $\mathscr{R}_2 \cap B$ is non-empty. Pick any $\succsim_2 \in \mathscr{R}_2 \cap B$. Note that since $B \subset \mathscr{R}_1$, it follows that $\succsim_2 \in \mathscr{R}_1$. For each $i \geq 3$, pick any

$\succsim_i \in \mathscr{R}_i$. The profile $(\succsim_2, \succsim_2, \succsim_3, \dots, \succsim_n) \in \mathscr{R}$ clearly violates Independent Prospects.

9 Conclusion

This paper's main purpose has been to develop a simple and transparent axiomatization of relative utilitarianism using the concept of the revealed marginal rates of substitution. We have done so considering a subset of the set of all preference profiles. We could try to extend our result by considering the complete set of all profiles of von Neumann-Morgenstern preferences by requiring continuity of the marginal rates of substitution with respect to the topology for the domain introduced in the previous section. We would then obtain that for all profiles for which the social welfare function reveals a marginal rate of substitution between two agents i and j these two agents must have the same weight in the social welfare function. Nothing would follow if the social welfare function does not reveal a marginal rate of substitution, a possibility that we discussed in Section 3. When the social welfare function does not reveal any marginal rate of substitution for some pair of agents i and j (this can only be true for profiles that violate the Independent Prospects Condition) then our approach does not have any implications for the relative weight of i and j in the social welfare function. It seems natural that in such a case an approach based on revealed marginal rates of substitution does not make any predictions about how the social planner would choose.

In this paper we have made strong assumptions regarding the number of individuals and the number of alternatives. It appears worthwhile to investigate the implications of our axioms on domains that do not satisfy these assumptions. Further future work can include the investigation of the consequences of alternative axioms in our framework.

References

[1] Kenneth J. Arrow (1951), *Social Choice and Individual Values*, New York: Wiley.

[2] Alan Baker (1975), *Transcendental Number Theory*, Cambridge, England: Cambridge University Press.

[3] Abram S. Besicovitch (1940), On the Linear Independence of Fractional Powers of Integers, *Journal of the London Mathematical Society* 15, 3-6.

[4] Ken Binmore (2009), *Rational Decisions*, Princeton: Princeton University Press.

[5] Tilman Börgers and Yan-Min Choo (2017), A Counterexample to Dhillon (1998), *Social Choice and Welfare* 48, 837-843.

[6] Graciela Chichilnisky (1985), Von Neumann-Morgenstern Utilities and Cardinal Preferences, *Mathematics of Operations Research* 4, 633-641.

[7] Claude d'Aspremont and Louis Gevers (1977), Equity and the Informational Basis of Collective Choice, *Review of Economic Studies* 44, 199-209.

[8] Claude d'Aspremont and Louis Gevers (2002), Social Welfare Functionals and Interpersonal Comparability, in: Kenneth J. Arrow, Amartya Sen and Kotaro Suzumura (editors), *Handbook of Social Choice and Welfare*, Volume 1, Amsterdam: North Holland, 459-541.

[9] Amrita Dhillon (1998), Extended Pareto Rules and Relative Utilitarianism, *Social Choice and Welfare* 15, 521-542.

[10] Amrita Dhillon and Jean-François Mertens (1997), An Impossibility Theorem With von Neumann-Morgenstern Preferences, *Economics Letters* 56, 305-309.

[11] Amrita Dhillon and Jean-François Mertens (1999), Relative Utilitarianism, *Econometrica* 67, 471-498.

[12] Peter C. Fishburn (1969), Preferences, Summation, and Social Welfare Functions, *Management Science* 16, 179-186.

[13] Peter C. Fishburn (1973), Summation Social Choice Functions, *Econometrica* 41, 1183-1196.

[14] Peter C. Fishburn (1984), On Harsanyi's Utilitarian Cardinal Welfare Theorem, *Theory and Decision* 17, 21-28.

[15] Marcus Fleming (1952), A Cardinal Concept of Welfare, *The Quarterly Journal of Economics* 66, 366-384.

[16] Marcus Fleming (1957), Cardinal Welfare and Individualistic Ethics: A Comment, *Journal of Political Economy* 65, 355-357.

[17] John Harsanyi (1953), Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking, *Journal of Political Economy* 61, 434-435.

[18] John Harsanyi (1955), Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility, *Journal of Political Economy* 63, 309-321.

[19] John Harsanyi (1977), *Rational Behavior and Bargaining Equilibrium in Games and Social Situations,* Cambridge: Cambridge University Press, 1977.

[20] Aanund Hylland (1980), Aggregation Procedure for Cardinal Preferences: A Comment, *Econometrica* 48, 539-542.

[21] Ehud Kalai and David Schmeidler (1977), Aggregation Procedure for Cardinal Preferences: A Formulation and Proof of Samuelson's Impossibility Conjecture, *Econometrica* 45, 1431-1438.

[22] Edi Karni (1998), Impartiality: Definition and Representation, *Econometrica* 66, 1405-1415.

[23] Eric Maskin (1978), A Theorem on Utilitarianism, *Review of Economic Studies* 45, 93-96.

[24] Philippe Mongin (1994), Harsanyi's Aggregation Theorem: Multi-Profile Version and Unsettled Questions, *Social Choice and Welfare* 11, 331-354.

[25] Uzi Segal (2000), Let's Agree That All Dictatorships Are Equally Bad, *Journal of Political Economy* 108, 569-589.

[26] Amartya Sen (1970), *Collective Choice and Social Welfare*, San Francisco: Holden-Day, Inc.

[27] John A. Weymark (1991), A Reconsideration of the Harsanyi-Sen Debate on Utilitarianism, in: Jon Elster and John E. Roemer (editors), *Interpersonal Comparisons of Well-Being*, Cambridge etc.: Cambridge University Press, 255-320.

[28] John A. Weymark (1994), Harsanyi's Social Aggregation Theorem With Alternative Pareto Principles, in: Wolfgang Eichhorn (editor), *Models and Measurement of Welfare and Inequality*, Heidelberg: Springer Verlag, 869-887.

Appendix: Proof of Proposition 3

It will be convenient to sometimes write the set of alternatives as $A = \{1, 2, \ldots, m\}$ and at other times write it as $A = \{a_1, a_2, \ldots, a_m\}$. We first construct for each agent i the set $U_i \subset [0,1]^m$ of possible von Neumann-Morgenstern utility representations of her preferences. We then define that agent's set of possible preference relations by:

$$\mathscr{R}_i = \{\succsim_i \in \mathcal{R} | \exists u_i \in U_i : u(\succsim_i) = u_i\}. \tag{26}$$

Let $p : \mathbb{N} \to \mathbb{R}$ be the map that assigns to every $x \in \mathbb{N}$ the square root of the x^{th} prime number (so $p(1) = \sqrt{2}, p(2) = \sqrt{3}, p(3) = \sqrt{5}, p(4) = \sqrt{7}$, etc.). Define, for each $i \in N$ and each $a \in A$:

$$T_{i,a} = \{qe^{p(im+a)} | q \in \mathbb{Q}\} \cap (0, 1). \tag{27}$$

Now define U_i to be the set of vectors $u_i = (u_{i,1}, u_{i,2}, \ldots, u_{i,m})$ with these properties: Each $u_{i,a} \in T_{i,a} \cup \{0, 1\}$; and the number of entries in u_i which read 1 is one, two, or three, while the number which read 0 is one or two. This completes our construction, for each $i \in N$, of the set U_i and hence also the set $\mathscr{R}_i$. Define $\mathscr{R} = \mathscr{R}_1 \times \mathscr{R}_2 \times \cdots \times \mathscr{R}_n$. We now verify that $\mathscr{R}$ is a domain that satisfies the three assumptions.

Assumption 1: The first sentence is obviously satisfied. It remains to prove that every profile $\succsim = (\succsim_1, \succsim_2, \ldots, \succsim_n) \in \mathscr{R}$ satisfies Independent Prospects. The preference relations $\succsim_1, \succsim_2, \ldots, \succsim_n$ correspond to the normalized utility vectors $u_1, u_2, \ldots, u_n$. Use these vectors to form this

$m \times (n+1)$ matrix:

$$\begin{pmatrix} | & | & \dots & | & 1 \\ | & | & \dots & | & \vdots \\ u_1 & u_2 & \dots & u_n & 1 \\ | & | & \dots & | & \vdots \\ | & | & \dots & | & 1 \end{pmatrix}_{m\times(n+1)} \tag{28}$$

By construction, each u_i has at most 5 entries that are not elements of some $T_{i,a}$. Hence, the above matrix has at most $5n$ rows that contain an entry that reads either 0 or 1. By the assumption that $m > 6n$, this means that there are at least $n+1$ rows of the above matrix whose entries are all members of some $T_{i,a}$. Take any such $n+1$ rows to form this $(n+1)\times(n+1)$ sub-matrix:

$$\begin{pmatrix} | & | & \dots & | & 1 \\ | & | & \dots & | & \vdots \\ v_1 & v_2 & \dots & v_n & 1 \\ | & | & \dots & | & \vdots \\ | & | & \dots & | & 1 \end{pmatrix}_{(n+1)\times(n+1)} \tag{29}$$

We claim that the determinant of this sub-matrix is non-zero. The argument is as follows. This determinant can be expressed as a non-trivial rational polynomial in $e^{p(1)}, e^{p(2)}, \dots, e^{p(n\cdot m)}$. The Lindemann-Weierstrass Theorem (Theorem 1.4 in Baker [2]) says that such a polynomial is non-zero if the numbers $p(1), p(2), \dots, p(n\cdot m)$ are algebraic and linearly independent over $\mathbb{Q}$. That they are algebraic is obvious. That they are linearly independent over $\mathbb{Q}$ is shown in Theorem 2 in Besicovitch [3].

Because the determinant of the above sub-matrix is non-zero, we can find for every i, some non-zero vector $r_i \in \mathbb{R}^{n+1}$ such that $r_i \cdot v_i \neq 0$, $r_i \cdot \mathbf{1} = 0$, and for all $j \neq i$, $r_i \cdot v_j = 0$. Pick any $\hat{p}_i \in \mathbb{R}^{n+1}_{+}\backslash\{\mathbf{0}\}$ such that $\hat{p}_i \gg r_i$. Let $\hat{q}_i = \hat{p}_i - r_i$. Observe that $\hat{q}_i \in \mathbb{R}^{n+1}_{+}\backslash\{\mathbf{0}\}$ and that $\hat{q}_i \cdot \mathbf{1} = (\hat{p}_i - r_i)\cdot \mathbf{1} = \hat{p}_i \cdot \mathbf{1} - r_i \cdot \mathbf{1} = \hat{p}_i \cdot \mathbf{1}$. Now divide both $\hat{p}_i$ and $\hat{q}_i$ by $\hat{p}_i \cdot \mathbf{1}$ to get $\tilde{p}_i = \hat{p}_i/(\hat{p}_i \cdot \mathbf{1})$ and $\tilde{q}_i = \hat{q}_i/(\hat{p}_i \cdot \mathbf{1})$. Let p_i be the lottery that assigns to the $n+1$ alternatives (that were involved in forming the sub-matrix) probability weights as per the probability vector $\tilde{p}_i$ and assigns to all other alternatives probability weight 0. Analogously, let q_i be the lottery that assigns to the $n+1$ alternatives probability weights as per the probability vector $\tilde{q}_i$ and assigns to all other alternatives probability weight 0. We now verify that the lotteries p_i, q_i satisfy the conditions in Independent Prospects.

For any $j \neq i$,

$$\begin{aligned} p_i \cdot u_j - q_i \cdot u_j &= \tilde{p}_i \cdot v_j - \tilde{q}_i \cdot v_j \\ &= (\tilde{p}_i - \tilde{q}_i) \cdot v_j = (\hat{p}_i - \hat{q}_i) \cdot v_j / \hat{p}_i \cdot \mathbf{1} \\ &= r_i \cdot v_j / \hat{p}_i \cdot \mathbf{1} = 0 / \hat{p}_i \cdot \mathbf{1} = 0. \end{aligned} \tag{30}$$

So $p_i \cdot u_j = q_i \cdot u_j$, that is to say, $p_i \sim_j q_i$. On the other hand,

$$\begin{aligned} p_i \cdot u_i - q_i \cdot u_i &= \tilde{p}_i \cdot v_i - \tilde{q}_i \cdot v_i = (\tilde{p}_i - \tilde{q}_i) \cdot v_i \\ &= (\hat{p}_i - \hat{q}_i) \cdot v_i / \hat{p}_i \cdot \mathbf{1} = r_i \cdot v_i / \hat{p}_i \cdot \mathbf{1} \neq 0. \end{aligned} \tag{31}$$

So $p_i \cdot u_i \neq q_i \cdot u_i$, that is to say, $p_i \not\sim_i q_i$.

Assumption 2: (ii) is obviously satisfied. We prove (i). Let $\succsim_i = \succsim_i^1, \hat{\succsim}_i = \succsim_i^K \in \mathscr{R}_i$. If $\succsim_i^1$ ranks more than one alternative top, then pick any top alternative (call it a_1). Let $\succsim_i^2 \in \mathscr{R}_i$ be the preference relation that assigns to the alternative a_1 some Bernoulli utility from the set T_{i,a_1} and assigns to all other alternatives the same Bernoulli utility as did $\succsim_i^1$. Repeat this procedure as many times as is possible, to arrive at some $\succsim_i^a \in \mathscr{R}_i$ that ranks exactly one alternative top (call this alternative a_a). Through an analogous procedure, we can arrive at some $\succsim_i^b \in \mathscr{R}_i$ that ranks exactly one alternative bottom (call it a_b) and still ranks a_a as the only top alternative.

Now pick any alternative a_{b+1} that isn't a_b or a_a. Let $\succsim_i^{b+1} \in \mathscr{R}_i$ be the preference relation that ranks a_{b+1} top and assigns to all other alternatives the same Bernoulli utility as did $\succsim_i^b$. So now $\succsim_i^{b+1}$ ranks a_a and a_{b+1} as the only top alternatives and a_b as the only bottom alternative.

Now pick any alternative a_c that is ranked bottom by $\hat{\succsim}_i$. If $a_c = a_b$ (meaning that $\succsim_i^{b+1}$ already ranked a_c bottom), then simply let $\succsim_i^c = \succsim_i^{b+1}$. Otherwise, let $\succsim_i^{b+2} \in \mathscr{R}_i$ be the preference relation that ranks a_c bottom and assigns to all other alternatives the same Bernoulli utility as did $\succsim_i^{b+1}$. So now $\succsim_i^{b+2}$ ranks as top only a_{b+1} (and also a_a, if $a_a \neq a_c$) as the only top alternatives and as bottom only a_b and a_c. Next, let $\succsim_i^c \in \mathscr{R}_i$ be the preference relation that assigns to the alternative a_b some Bernoulli utility from the set $T_{i,a_b}$and assigns to all other alternatives the same Bernoulli utility as did $\succsim_i^{b+2}$. So now $\succsim_i^c$ ranks a_c as the only bottom alternative.

Next pick any alternative a_d that is ranked top by $\hat{\succsim}_i$. Through similar steps, we can get from $\succsim_i^c$ to some $\succsim_i^d$ that continues to rank a_c as the only bottom, but now also ranks a_d as the only top.

Now construct $\succsim_i^{d+1}, \succsim_i^{d+2}, \ldots, \succsim_i^{d+m}$ as follows: Let $\succsim_i^{d+s} \in \mathscr{R}_i$ be the preference relation that assigns to the alternative a_s the same Bernoulli utility as does $\hat{\succsim}_i$ and assigns to all other alternatives the same Bernoulli utility as did $\succsim_i^{d+s-1}$. We have that $\succsim_i^{d+m} = \succsim_i^K = \hat{\succsim}_i$. This completes the construction of a sequence of single modifications connecting $\succsim_i$ to $\succsim_i^K$.

Assumption 3: Each of the constructed sets $T_{i,a}$ is obviously dense in $[0, 1]$. So D_i is dense in $[0, 1]^m$. Thus $\mathscr{R}_i$ is dense in $\mathcal{R}$.