Social Choice and Welfare (2024) 63:1–18
https://doi.org/10.1007/s00355-024-01518-1

ORIGINAL PAPER

Impartiality and relative utilitarianism

Edi Karni[1] · John A. Weymark[2]

Received: 3 August 2023 / Accepted: 26 February 2024 / Published online: 30 March 2024

Abstract
A novel axiomatization of relative utilitarianism is provided using the single-profile setting used in Harsanyi's Social Aggregation Theorem. Harsanyi's axioms are supplemented with an impartiality axiom that requires social alternative lotteries p and q to be socially indifferent when (i) two individuals have conflicting preferences for them and everybody else is indifferent and (ii) the concerned individuals' strengths of preference for p over q have the same magnitude. This axiomatization shows that equality of the social weights can be obtained in a single-profile setting and that no interprofile condition is needed to obtain profile-independent weights in a multi-profile setting.

1 Introduction

Classical utilitarianism ranks social alternatives by the sum of their utilities.[1] Relative utilitarianism normalizes utilities before summing so that each individual has a maximum utility of 1 and a minimum utility of 0. Axiomatizations of relative utilitarianism have been provided by Karni (1998), Dhillon (1998), Dhillon and Mertens (1999), Segal (2000), Börgers and Choo (2017b), and Brandl (2021).[2] We offer a novel axiomatization of this social welfare criterion by supplementing Harsanyi's axioms with an impartiality axiom that applies when comparing two lotteries for which two

[1] Classical utilitarianism is neutral about what conception of well-being the utilities measure.
[2] Börgers and Choo (2017a) have provided a counterexample to one of the results that Dhillon (1998) uses to characterize relative utilitarianism. It is an open question if her axiomatization is correct.

✉ John A. Weymark
john.weymark@vanderbilt.edu

Edi Karni
karni@jhu.edu

[1] Department of Economics, Johns Hopkins University, 3400 N. Charles Street, 544E Wyman Bldg., Baltimore, MD 21218, USA

[2] Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235-1819, USA

individuals have conflicting preferences of equal strength and for which everybody else is indifferent between them.

Our characterization of relative utilitarianism builds on Harsanyi's Social Aggregation Theorem (Harsanyi 1955). Harsanyi considers a single profile of individual preferences and a social preference relation on the set of lotteries generated by a finite set of social alternatives. We interpret the social preference as being that of a social observer but it could also be the ethical preferences of some individual. Harsanyi shows that if (i) the individual and social preference relations satisfy the axioms of expected utility theory and are represented by von Neumann–Morgenstern utility functions (von Neumann and Morgenstern 1944) and (ii) they are related by a Pareto condition, then the social alternative lotteries are socially ranked by a weighted sum of the individual utilities obtained with them. With the Strong Pareto version of the Pareto condition, the welfare weights can be chosen to be positive. If, furthermore, the individuals' utilities can be varied independently, a property known as Independent Prospects (Weymark 1991), the weights are unique.

Harsanyi has argued that his Social Aggregation Theorem provides a decision-theoretic foundation for utilitarianism. This inference has been disputed by Sen (1976) using an argument later formalized by Weymark (1991). For utilitarianism to be a meaningful doctrine, it must be possible to make interpersonal comparisons of utility gains and losses. The axioms of expected utility theory only place restrictions on a preference relation that ranks pairs of lotteries and, therefore, expected utility theory is ordinal. Sen and Weymark note that while preferences satisfying the axioms of expected utility theory *may* be represented by von Neumann–Morgenstern utility functions, they need not be—any increasing transforms of such functions represent the preferences equally well.[3] If the individual preferences are represented by nonlinear transforms of von Neumann–Morgenstern utility functions, then the social alternative lotteries are not socially ranked by a weighted sum of utilities and, consequently, the utilitarian interpretation of Harsanyi's Theorem is unjustified.

As in Harsanyi's Social Aggregation Theorem, we assume that (i) the set of alternatives is the set of lotteries on a finite set of social alternatives, (ii) there is a single profile of individual preferences and a social preference on this set, and (iii) all of these preferences satisfy the expected utility axioms. Harsanyi's axioms are supplemented by an impartiality axiom that is concerned with how the conflicting interests of two individuals are adjudicated in some *two-person situations*. In such a situation, some individual j prefers social alternative lottery p to lottery q, some other individual k prefers q to p, and everybody else is indifferent between p and q. Our impartiality axiom requires p to be socially indifferent to q in a two-person situation when the strength of preference for p over q is the same in absolute value for the two concerned individuals.

We use an individual's 0-1 normalized von Neumann–Morgenstern utility function to measure his strength of preference for any social alternative lottery p relative to any other social alternative lottery q. We make a normative judgment that the strengths of preference as so measured provide the appropriate way of comparing utility gains and

[3] See also Luce and Raiffa (1957, p. 32) and Montesano (1985).

losses both intrapersonally and interpersonally. It is this normative assumption that allows us to circumvent the Sen–Weymark critique.

One can distinguish between three questions: (i) Is strength of preference a meaningful concept? (ii) Can strength of preference be measured as a difference in an individual's 0-1 normalized von Neumann–Morgenstern utility function? (iii) Can one compare different individuals' strengths of preference?[4] In each case, we believe that the answer is yes. However, we do not claim that that there is a unique way to measure strength of preference.

In our axiomatization of relative utilitarianism, we treat Independent Prospects and the requirement that the individual and social preferences satisfy the expected utility axioms as maintained assumptions. We show that the social alternative lotteries are socially ranked according to the relative utilitarian criterion using 0-1 normalized von Neumann–Morgenstern utility functions if and only if Strong Pareto and our impartiality axiom are satisfied.

Our axiomatization implies that (i) it is possible to establish that the social weights are equal without employing a multi-profile setting and (ii) if it is applied profile-by-profile in such a setting, no interprofile condition is needed to obtain profile-independent weights. These conclusions run counter to claims made by Mongin (1994) and Mongin and d'Aspremont (1998).

A response to the Sen–Weymark critique is that interpersonal utility comparisons are revealed by the choice behavior of the social observer. A revealed preference interpretation of a multi-profile version of Harsanyi's Social Aggregation Theorem is advocated by Binmore (2009, Chap. 4) and used by Börgers and Choo (2017b) to axiomatize relative utilitarianism.[5] In this approach, the social weights that are used to aggregate the individual utilities reveal how the social observer makes interpersonal comparisons of utility gains and losses. Specifically, the ratio of two individuals' weights reveals how a utility difference for one of them is converted into a utility difference for the other. However, what these weights are depends on which representations of the individual utility functions are used. To justify particular representations requires further argumentation, such as that provided by Börgers and Choo.

Of the existing axiomatizations of relative utilitarianism, only Karni (1998) uses the single-profile framework employed here and by Harsanyi (1955). In contrast, Dhillon (1998), Dhillon and Mertens (1999), and Börgers and Choo (2017b) consider a multi-profile problem in which a social preference over the set of social alternative lotteries must be determined for each profile of individual preferences in some domain.[6] Brandl (2021) also uses a multi-profile approach but models uncertainty as in Savage (1954). Segal (2000) considers how to socially choose among lotteries over possible divisions of a bundle of resources as the quantities of these resources are varied. Below, we

[4] The value of separating these three issues was suggested to us by a referee.

[5] This approach is implicit in Mongin (1994). Using a multi-profile welfarist approach, Mongin argues that if social preferences satisfy expected utility theory's independence condition, then interpersonal comparisons of utility gains and losses are implicitly being made. See also Mongin and d'Aspremont (1998, Sec. 5).

[6] Börgers and Choo (2017b) provide a good introduction to the contributions of Dhillon (1998) and Dhillon and Mertens (1999).

discuss how Börgers and Choo's revealed preference approach to making comparisons of utility gains and losses differs from our own.[7]

Assumptions about how to reconcile conflicting interests in two-person situations have been previously used by Karni (1998, 2003) and Raschka (2023) to help axiomatize various utilitarian principles. We defer a discussion of their impartiality axioms until after we have formally introduced our own.

The plan of this article is as follows. In Sect. 2, we describe the model and present Harsanyi's Social Aggregation Theorem. In Sect. 3, we introduce our impartiality axiom. Our axiomatization of relative utilitarianism is presented in Sect. 4. In Sect. 5, we compare our impartiality axiom with other criteria that have been used to resolve conflicting interests. In Sect. 6, we comment on Börgers and Choo's revealed preference approach. We offer some concluding remarks in Sect. 7.

2 Harsanyi's social aggregation theorem

In this section, we introduce our model and provide a formal statement of a Strong Pareto version of Harsanyi's (1955) Social Aggregation Theorem.

The set of individuals is $N = \{1, \ldots, n\}$, where $n \geq 2$. The finite set of *social alternatives* (outcomes) is $X = \{x_1, \ldots, x_m\}$, where $m \geq 2$. Let $\Delta(X) = \{p \in \mathbb{R}^m \mid \sum_{x \in X} p(x) = 1, p(x) \geq 0, x \in X\}$ be the set of *social alternative lotteries*. We denote by δ_x the social alternative lottery that assigns the unit probability mass to the alternative x. For each $i \in N$, let $\succeq_i$ be a binary relation on $\Delta(X)$ representing the preference ordering of individual i. Let $\succeq_0$ be a binary relation on $\Delta(X)$ representing the social preference ordering and denote by N^0 the union $N \cup \{0\}$. For every preference relation $\succeq_i$, $i \in N^0$, we define the strict preference relation, $\succ_i$, and the indifference relation, $\sim_i$, as usual. A preference relation $\succeq_i$ is *degenerate* if $\succ_i$ is empty and it is *nondegenerate* otherwise.

Harsanyi assumes that both the individual and social preferences satisfy the axioms of expected utility theory (ordering, continuity, and independence).

Axiom A.1 *(Expected Utility) For each $i \in N^0$, $\succeq_i$ satisfies the axioms of expected utility theory.*

Because the preference relation $\succeq_i, i \in N^0$, is a continuous ordering, it can be represented by a *utility function*. That is, for each $i \in N^0$, there exists a function $v_i \colon \Delta(X) \to \mathbb{R}$ such that

$$p \succeq_i q \leftrightarrow v_i(p) \geq v_i(q), \quad \text{for all } p, q \in \Delta(X). \tag{1}$$

As von Neumann and Morgenstern (1944) have shown, if Axiom A.1 is satisfied, then v_i can be chosen so that

[7] Sprumont (2013) axiomatizes the leximin rule defined using 0-1 normalized von Neumann–Morgenstern utility representations. In his problem, individuals have expected utility preferences over social alternative lotteries whose outcomes are allowed to vary.

$$v_i(p) = \sum_{l=1}^{m} p_l v_i(\delta_{x_l}), \quad \text{for all } p \in \Delta(X). \tag{2}$$

Identifying the lottery δ_{x_l} with the alternative x_l, for each $i \in N^0$, we can define a function $u_i : X \to \mathbb{R}$ so that $u_i(x_l) = v_i(\delta_{x_l})$. With this notation, (2) may be rewritten as

$$v_i(p) = \sum_{l=1}^{m} p_l u_i(x_l), \quad \text{for all } p \in \Delta(X). \tag{3}$$

Thus, the utility of a social alternative lottery is the expected value of the utility obtained with the social alternative that is realized once the uncertainty is resolved. The functions v_i and u_i are each called a *von Neumann–Morgenstern utility function* when utilities have the expected utility form given in (2) and (3). If the preference relation $\succeq_i$ is nondegenerate, the choice of v_i and u_i in (2) and (3) is unique up to an increasing affine transform.[8]

Harsanyi requires the social preference to satisfy a Pareto principle. We consider a strong form of this principle.

Axiom A.2 *(Strong Pareto) For all $p, q \in \Delta(X)$, $p \succeq_i q$ for all $i \in N$ implies $p \succeq_0 q$, and if, in addition, $p \succ_i q$ for some $i \in N$, then $p \succ_0 q$.*

In the proof of his Social Aggregation Theorem, Harsanyi implicitly assumed that the individual preference relations are distinct in the following sense. For each individual, there is a pair of social alternative lotteries between which he is not indifferent but for which everybody else is. This condition on the profile of preference relations $\{\succeq_i\}_{i \in N}$ is called *Independent Prospects*.

Axiom A.3 *(Independent Prospects) For all $i \in N$, there exist $p^i, q^i \in \Delta(X)$ such that $p^i \sim_j q^i$ for all $j \in N \backslash \{i\}$ and $\neg(p^i \sim_i q^i)$.*

Independent Prospects implies that each of the individual preference relations $\succeq_i$, $i \in N$, is nondegenerate. Note that this axiom requires m to be larger than n.

Harsanyi's Social Aggregation Theorem in its Strong Pareto form shows that if Axioms A.1–A.3 are satisfied, then for any von Neumann–Morgenstern utility functions chosen to represent the individual and social preference relations, the social utility function must be a positive weighted sum of the individual utility functions modulo the addition of a constant term.

Theorem 1 (Harsanyi's Theorem) *Suppose that $\{\succeq_i\}$, $i \in N$, are preference relations on $\Delta(X)$ that satisfy Axiom A.1 and that these relations jointly satisfy Axiom A.3. Further suppose that $\succeq_0$ is a preference relation on $\Delta(X)$ that satisfies Axiom A.1 and that $\{\succeq_i\}$, $i \in N^0$, jointly satisfy Axiom A.2. If $\succeq_i$, $i \in N^0$, is represented by the von Neumann–Morgenstern utility function $v_i : \Delta(X) \to \mathbb{R}$, then there exist unique social weights $w_i > 0$, $i \in N$, and a unique scalar c such that*

$$v_0(p) = \sum_{i=1}^{n} w_i v_i(p) + c, \quad \textit{for all } p \in \Delta(X). \tag{4}$$

[8] A function $f : \mathbb{R} \to \mathbb{R}$ is an *increasing affine transform* if $f(t) = a + bt$ with $b > 0$.

Alternative Pareto conditions have different implications about the signs of the welfare weights in (4). If Independent Prospects is omitted from the assumptions of Theorem 1, then the form of the aggregation equation in (4) is unchanged but the parameters in it are then not unique and the welfare weights need not all be positive.[9]

3 Impartiality

It may be difficult to determine whether any particular way of ranking two social alternative lotteries accords with our ethical intuitions when there are many individuals who are not indifferent between them and the concerned individuals do not agree about how they should be ranked. There may be more confidence in a social evaluation about how to rank two social alternative lotteries if there are only two concerned individuals. For this reason, the impartiality axiom introduced below only applies to such two-person situations, which are formally defined as follows.[10]

Definition 1 For all distinct $j, k \in N$ and all distinct $p, q \in \Delta(X)$, (j, k, p, q) is a *two-person situation* if $p \succ_j q$, $q \succ_k p$, and $p \sim_i q$ for all $i \in N \backslash \{j, k\}$.[11]

Our impartiality axiom takes account of the strength of the conflicting interests of the two concerned individuals in a two-person situation and, therefore, supplements the ordinal information about the individual preferences with cardinal information about strengths of preference.[12] We assume that the measure of an individual's strength of preference defined below also provides the normatively significant measure of an individual's utility gain or loss for the purpose of making intrapersonal and interpersonal utility comparisons. Before defining our strength of preference measure, we first need to introduce some further notation.

Suppose that $\succeq_i$ is a nondegenerate expected utility preference for all $i \in N$. For each $i \in N$, let $\hat{x}_i, \check{x}_i \in X$ be, respectively, $\succeq_i$-best and $\succeq_i$-worst elements of X. Formally, $\delta_{\hat{x}_i} \succeq_i p \succeq_i \delta_{\check{x}_i}$ for all $p \in \Delta(X)$. Because X is finite, $\hat{x}_i$ and $\check{x}_i$ exist. Moreover, because $\succeq_i$ is nondegenerate, $\hat{x}_i \neq \check{x}_i$. If $\hat{x}_i$ is nonunique, we choose $\hat{x}_i$ from the set of $\succeq_i$-best elements of X arbitrarily. The same is true for $\check{x}_i$.

For each $i \in N$, implicitly define a function $\Phi_i : \Delta(X) \to [0, 1]$ by

$$p \sim_i \left[\Phi_i(p)\delta_{\hat{x}_i} + (1 - \Phi_i(p))\,\delta_{\check{x}_i}\right]. \tag{5}$$

[9] For a discussion of different variants of Harsanyi's Social Aggregation Theorem, see Weymark (1991).

[10] A number of characterizations of social welfare orderings found in social choice theory are obtained by specifying the social ordering in all situations in which at most two people are not indifferent and then using some axioms that have widespread support to infer the complete social ordering. See Bossert and Weymark (2004, Sec. 12).

[11] If $n = 2$, then the requirement that $p \sim_i q$ for all $i \in N \backslash \{j, k\}$ is vacuous.

[12] With an ordinal preference, it may be possible to make some inferences concerning preference strengths without invoking non-ordinal information. For example, suppose that $p \succ_i q \succ_i r \succ_i s$. Then, one can infer that i's strength of preference for p over s is larger than it is for q over r. However, the circumstances for which this and and related inferences can be made are too limited for our purpose. See Baccelli (2024) for an illuminating investigation of ordinal utility differences.

Because $\succeq_i$ is nondegenerate and satisfies the axioms of expected utility theory, for each $p \in \Delta(X)$, there is a unique probability mixture between $\delta_{\hat{x}_i}$ and $\delta_{\check{x}_i}$ that i regards as being indifferent to p. Hence, Φ_i is well-defined.

Consider two social alternative lotteries $\bar{p}, \bar{q} \in \Delta(X)$ that only put positive probability on i's best and worst outcomes (i.e., $\bar{p} = \bar{p}_{\hat{x}_i}\delta_{\hat{x}_i} + (1 - \bar{p}_{\hat{x}_i})\delta_{\check{x}_i}$ and $\bar{q} = \bar{q}_{\hat{x}_i}\delta_{\hat{x}_i} + (1 - \bar{q}_{\hat{x}_i})\delta_{\check{x}_i}$). In this case, we regard the difference $\bar{p}_{\hat{x}_i} - \bar{q}_{\hat{x}_i}$ as a measure of i's strength of preference for $\bar{p}$ over $\bar{q}$. That is, this strength of preference is measured by the amount by which the likelihood of achieving i's best outcome is changed when $\bar{q}$ is replaced by $\bar{p}$. If $\bar{q} \succ_i \bar{p}$, the strength of preference for $\bar{p}$ over $\bar{q}$ is negative. The value $\bar{p}_{\hat{x}_i} - \bar{q}_{\hat{x}_i}$ can also be given a willingness-to-pay interpretation. It is the probability premium that individual i is willing to pay for being allowed to choose $\bar{p}$ instead of $\bar{q}$, where the premium is expressed in terms of the probability of obtaining i's most preferred outcome.

In the preceding discussion, we have considered the comparison of two social alternative lotteries in which only $\hat{x}_i$ and $\check{x}_i$ are given positive probability. We can also regard $\bar{p}_{\hat{x}_i} - \bar{q}_{\hat{x}_i}$ as being the strength of preference for p over q for any two social alternative lotteries p and q that are indifferent to $\bar{p}$ and $\bar{q}$, respectively. Noting that $\Phi_i(\bar{p}) = \bar{p}_{\hat{x}_i}$ and $\Phi_i(\bar{q}) = \bar{q}_{\hat{x}_i}$, this leads to the following definition.

Definition 2 For all $i \in N$ and all $p, q \in \Delta(X)$, i's *strength of preference for p over q* is $\Phi_i(p) - \Phi_i(q)$.

This way of defining an individual's strength of preference has an affinity with a proposal made by von Neumann and Morgenstern (1944, p. 18). They consider three social alternatives x_1, x_2, x_3 for which $\delta_{x_1} \succ_i \delta_{x_2} \succ_i \delta_{x_3}$ for some individual i. Letting α be such that $[\alpha\delta_{x_1} + (1-\alpha)\delta_{x_3}] \sim_i \delta_{x_2}$, they suggest that it is "plausible" to regard α as providing a numerical measure for comparing i's preference for x_1 over x_2 to his preference for x_2 over x_3.

We assume that $\Phi_i(p) - \Phi_i(q)$ not only measures i's strength of preference for p over q, but also that it measures his utility gain or loss for the transition from q to p. We are able to identify an individual's strength of preference with a utility difference because the function Φ_i is the unique von Neumann–Morgenstern utility function that represents $\succeq_i$ for which $\Phi_i(\delta_{\hat{x}_i}) = 1$ and $\Phi_i(\delta_{\check{x}_i}) = 0$.

An alternative way of measuring utility differences was introduced by Alt (1936, 1971). For Alt, utility differences are defined by a representation of a quaternary relation $\succeq^*$ over pairs of alternatives for which $(a, b) \succeq^* (c, d)$ is interpreted as meaning that the transition from b to a is weakly preferred to the transition d to c. Alt identified conditions under which the utility for the transition from b to a can be written as $U(a) - U(b)$ for some utility function U representing a preference on the set of alternatives. He further showed that the function U is unique up to an affine transform.[13] In contrast, here a preference for a transition between two alternatives is not a primitive of the model but is instead inferred from a preference on $\Delta(X)$ and our definition of strength of preference.

A utilitarian axiology requires that it be possible to compare utility gains and losses interpersonally. It does not follow from our assumption that an individual's strength of

[13] For histories of utility theory that focus on measurement issues, see Ellingsen (1994), Baccelli and Mongin (2016), and Moscati (2019).

preference for a pair of social alternative lotteries also measures his utility difference between them that the utility differences as so measured are interpersonally comparable. We assume that they are.[14] In other words, for the purpose of social evaluation, we are assuming that our measure of strength of preference provides the normatively significant way of computing utility gains and losses both intrapersonally and interpersonally. By making this assumption, we are adopting the position of Mongin (1994, p. 352), who argues that only a normative claim can establish that von Neumann–Morgenstern expected utility theory identifies the intensities of preferences that are relevant for social evaluation. Furthermore, we assume that our measure of strength of preference embodies the ethical views of the social observer about how differences in well-being are to be measured.[15] A consequence of our way of measuring strength of preference is that we are treating individuals *as if* they all have the same capacities for achieving well-being even if they do not. In this regard, we follow Robbins (1938, p. 637), who argues that "the postulate of equal capacity for satisfaction ... rest[s] upon ethical principle rather than scientific demonstration."

An implication of our identification of utility differences with strengths of preference is that for all $i, j \in N$, $\Phi_i(p) - \Phi_i(q)$ and $\Phi_j(p') - \Phi_j(q')$ are utility increments of equal magnitude when $\Phi_i(p) - \Phi_i(q) = \Phi_j(p') - \Phi_j(q')$. Thus, this approach to measuring utility differences can be regarded as the continuous analogue of the assumption in the finite case that utility differences between adjacent alternatives in a linear order represent constant utility increments intrapersonally that are of the same magnitude interpersonally.

Our approach can be contrasted with that of Edgeworth (1881). Edgeworth was a prominent utilitarian who used just-noticeable increments of pleasure to measure a unit of utility both intra- and interpersonally.[16] Ng (1975) and Argenziano and Gilboa (2019) have provided axiomatizations of weighted utilitarianism using just-noticeable differences as a basis for making comparisons of utility gains and losses.

We now turn to our impartiality axiom. Consider a two-person situation (j, k, p, q). Suppose that j's strength of preference for p over q is the same as k's strength of preference for q over p. This assumption is equivalent to requiring that $\Phi_j(p) - \Phi_j(q) = \Phi_k(q) - \Phi_k(p)$. Both of the differences in this equation are positive because (j, k, p, q) is a two-person situation. The kind of impartiality that we consider requires that p be socially indifferent to q in these circumstances.

Axiom A.4 *(Impartiality) For all distinct $j, k \in N$ and all distinct $p, q \in \Delta(X)$, if (j, k, p, q) is a two-person situation and if j's strength of preference for p over q is the same as k's strength of preference for q over p, then $p \sim_0 q$.*

[14] The strength of preference $\Phi_i(p) - \Phi_i(q)$ is a difference in probabilities and is therefore a dimensionless number. When this difference is reinterpreted as being a utility difference, this value is in units of i's utils. In order to compute a weighted sum of the utilities of different people, either the utilities must be measured using the same units or each person's weight must also be dimensioned (with the unit given by the inverse of the unit that that person's utility is measured in). We assume that the same units are being used. See Nebel (2022, 2021) for discussions of dimensioned and dimensionless quantities in social welfare analysis.

[15] For discussions of alternative bases for making comparative judgments about well-being, see Maniquet (2016) and Raschka (2023).

[16] For a brief introduction to Edgeworth's ways of measuring utility, see Moscati (2019, pp. 53–54).

Informally, if the interests of only two individuals conflict on the social alternative lotteries p and q and if the strengths of preference of the two concerned individuals are of equal magnitude but opposite in sign, then to treat them impartially requires p and q to be socially indifferent.

When it is assumed, as we do, that the functions Φ_i, $i \in N$, are 0-1 normalized von Neumann–Morgenstern utility functions and that they provide the basis for computing strengths of preference, our impartiality axiom can also be given a welfarist interpretation. In this interpretation, it is the fact that the utility gain for one person is equal in magnitude to the utility loss of a second person in a two-person situation that makes their circumstances equally meritorious and, therefore, that the two lotteries should be socially indifferent in order for these individuals to be treated impartially.

Our interpretation of Axiom A.4 as a principle of impartiality presupposes that the particular measure of strength of preference defined in Definition 2 is the appropriate one for determining equal merit in a two-person situation. However, our choice of how to measure strength of preference is not forced on us by the ordinal properties of the individual preferences, and so requires some justification. Why not, for example, measure person i's strength of preference for p over q by $\sqrt{\Phi_i(p) - \Phi_i(q)}$ and that of person j by $[\Phi_j(p) - \Phi_j(q)]^3$? With our identification of a utility difference with a strength of preference, the corresponding utility functions are 0-1 normalized, and so satisfy Robbins' ethical postulate that everybody is to be treated as if they have equal capacities for well-being.

A number of arguments can be advanced for adopting our measure of strength of preference when evaluating conflicting claims in two-person situations rather than any other that is ordinally equivalent to it. First, as in von Neumann and Morgenstern (1944), we could regard a probability difference as being a plausible or natural way of measuring preference intensity. Second, with a von Neumann–Morgenstern utility function, it is only necessary to know the utilities of sure outcomes in order to determine the expected utility of any lottery using (2) or (3). This simplifies the computation of the utilities that are to be aggregated.[17] Third, for $i, j \in N$ and $p, q \in \Delta(X)$, according to our definition of strength of preference, if $\Phi_i(p) - \Phi_i(q) = \Phi_j(q) - \Phi_j(p)$, then i and j have conflicting claims of equal merit. But then they also have conflicting claims of equal merit with the lotteries $[\alpha p + (1 - \alpha)q]$ and $[\beta p + (1 - \beta)q]$ for $0 < \alpha < \beta < 1$, which is not the case with any ordinal transforms of Φ_i and Φ_j (different from the identity transform) even if the 0-1 normalization is maintained. One could argue that this implication of our measure has normative appeal.

These three defenses of our way of measuring preference intensities all employ contentious arguments. However, in order to axiomatize relative utilitarianism by supplementing Harsanyi's axioms with an impartiality axiom that adjudicates between conflicting claims of equal merit (which is the objective here), it does not seem possible to employ any measure of preference intensity different from the one that we have proposed.

[17] These two arguments were suggested to us by a referee. Appeals to the simplicity and naturalness of the choice of a von Neumann–Morgenstern utility function to represent preferences that satisfy the expected utility axioms have been frequently made in the literature that has addressed the meaning and significance of von Neumann–Morgenstern expected utility theory. See Weymark (1991, 2005) for references and for some possible objections to using these desiderata when making social evaluations.

If there are no two-person situations involving j and k, then Impartiality does not apply to them. This would happen if either they have the same preferences or not everybody else is indifferent when j and k have conflicting preferences on a pair of lotteries. However, as we now show, if the profile of individual preferences satisfies Independent Prospects, then for any pair of distinct individuals, Impartiality is not vacuous.

Lemma *Suppose that $\{\succeq_i\}_{i\in N}$ is a profile of individual preference relations on $\Delta(X)$ each of which satisfies Axiom A.1 and that jointly satisfy Axiom A.3. Then, for every distinct $j, k \in N$, there exist distinct $p, q \in \Delta(X)$ such that $p \sim_i q$ for all $i \in N\backslash\{j, k\}$ and $\Phi_j(p) - \Phi_j(q) = \Phi_k(q) - \Phi_k(p) \neq 0$.*

Proof By Axiom A.1, each of the preference relations $\succeq_i$ has an expected utility representation Φ_i of the form defined implicitly in (5). Consider any distinct $j, k \in N$. By Axiom A.3, (i) there exist $p^j, q^j \in \Delta(X)$ such that $p^j \succ_j q^j$ and $p^j \sim_i q^j$ for all $i \neq j$ and (ii) there exist $p^k, q^k \in \Delta(X)$ such that $q^k \succ_k p^k$ and $p^k \sim_i q^k$ for all $i \neq k$. Thus, $\Phi_j(p^j) - \Phi_j(q^j) > 0$ and $\Phi_k(q^k) - \Phi_k(p^k) > 0$. There are three cases to consider.

Case 1: $\Phi_j(p^j) - \Phi_j(q^j) = \Phi_k(q^k) - \Phi_k(p^k)$. Let $p = 0.5p^j + 0.5p^k$ and $q = 0.5q^j + 0.5q^k$. Because Φ_j is a von Neumann–Morgenstern utility function, we have $\Phi_j(p) - \Phi_j(q) = [\Phi_j(0.5p^j + 0.5p^k)] - [\Phi_j(0.5q^j + 0.5q^k)] = [0.5\Phi_j(p^j) + 0.5\Phi_j(p^k)] - [0.5\Phi_j(q^j) + 0.5\Phi_j(q^k)] = 0.5[\Phi_j(p^j) - \Phi_j(q^j)]$, where the last equality follows because $p^k \sim_j q^k$. Similarly, $\Phi_k(q) - \Phi_k(p) = 0.5[\Phi_k(q^k) - \Phi_k(p^k)]$. Hence, $\Phi_j(p) - \Phi_j(q) = \Phi_k(q) - \Phi_k(p) > 0$. For all $i \in N\backslash\{j, k\}$, similar reasoning shows that $p \sim_i q$ because $\Phi_i(p^j) = \Phi_i(q^j)$ and $\Phi_i(p^k) = \Phi_i(q^k)$.

Case 2: $\Phi_j(p^j) - \Phi_j(q^j) > \Phi_k(q^k) - \Phi_k(p^k)$. By the continuity of Φ_j, there exists a $\lambda \in (0, 1)$ such that $\Phi_j(\bar{p}^j) - \Phi_j(q^j) = \Phi_k(q^k) - \Phi_k(p^k)$, where $\bar{p}^j = \lambda p^j + (1 - \lambda)q^j$. Because Φ_i is a von Neumann–Morgenstern utility function, everybody other than j is indifferent between $\bar{p}^j$ and q^j. Thus, the argument in Case 1 applies with $\bar{p}^j$ substituting for p^j.

Case 3: $\Phi_j(p^j) - \Phi_j(q^j) < \Phi_k(q^k) - \Phi_k(p^k)$. The proof of this case is the same as that of Case 2 with the roles of j and k reversed. □

4 A characterization of relative utilitarianism

Relative utilitarianism ranks social alternatives using the sum of 0-1 normalized utility functions. We specialize this definition to the special case in which the set of alternatives is the set of social alternative lotteries $\Delta(X)$ and the utility functions are the 0-1 normalized von Neumann–Morgenstern utility functions defined in (5).

Definition 3 For all $i \in N$, let Φ_i be the utility function representing the preference relation $\succeq_i$ on $\Delta(X)$ defined in (5). The social preference relation $\succeq_0$ is the *relative utilitarian* order for the utility functions $\{\Phi_i\}$, $i \in N$, if for all $p, q \in \Delta(X)$,

$$p \succeq_0 q \leftrightarrow \sum_{i=1}^{n} \Phi_i(p) \geq \sum_{i=1}^{n} \Phi_i(q). \tag{6}$$

The three axioms used in Harsanyi's Theorem (Theorem 1) do not characterize a unique social ordering. How the social alternative lotteries are ordered by the social aggregation equation (4) depends on the choice of of the von Neumann–Morgenstern utility functions used to represent the individual preferences and on the weights that are used to aggregate them. We now show that if these axioms are supplemented with our impartiality axiom, a unique social ordering is characterized: relative utilitarianism.

In our axiomatization of relative utilitarianism in the single-profile setting employed by Harsanyi (1955), we suppose that the individual and social preferences satisfy the expected utility axioms and that the individual preferences jointly satisfy Independent Prospects. We also suppose that the 0-1 normalized von Neumann–Morgenstern utility functions that represent the individual preference relations are used to compute the strengths of preference in our impartiality axiom. With these maintained assumptions, we show that Strong Pareto and Impartiality are satisfied if and only if the social preference is the relative utilitarian rule for these utility functions. Thus, it is only necessary to add Impartiality to the axioms in our version of Harsanyi's Social Aggregation Theorem in order to characterize relative utilitarianism.

Theorem 2 *Suppose that for all $i \in N$, $\succeq_i$ is a preference relation on $\Delta(X)$ that satisfies Axiom A.1 and that, for all $p, q \in \Delta(X)$, i's strength of preference for p over q is the utility difference $\Phi_i(p) - \Phi_i(q)$ for the 0-1 normalized von Neumann–Morgenstern utility function Φ_i that represents $\succeq_i$. Further suppose that the relations $\{\succeq_i\}$, $i \in N$, jointly satisfy Axiom A.3 and that $\succeq_0$ is a preference relation on $\Delta(X)$ that satisfies Axiom A.1. Then, the following conditions are equivalent:*

(i) *The relations $\{\succeq_i\}$, $i \in N^0$, jointly satisfy Axioms A.2 and A.4.*
(ii) *The relation $\succeq_0$ is the relative utilitarian order for the utility functions $\{\Phi_i\}$, $i \in N$.*

Proof It is straightforward to verify that (ii) implies (i), so we only consider the reverse implication.

Because Φ_i is a von Neumann–Morgenstern utility representation of $\succeq_i$ for all $i \in N$ and Axioms A.1–A.3 are satisfied, by Harsanyi's Theorem (Theorem 1), there exist unique positive weights w_i, $i \in N$, such that for all $p, q \in \Delta(X)$,

$$p \succeq_0 q \leftrightarrow \sum_{i=1}^{n} w_i \Phi_i(p) \geq \sum_{i=1}^{n} w_i \Phi_i(q). \tag{7}$$

Consider any distinct $j, k \in N$. Let $p, q \in \Delta(X)$ satisfy the assumptions of Axiom A.4 for these two individuals. By the Lemma, such p and q exist. By Axiom A.4, $p \sim_0 q$. Hence, by (7),

$$w_j[\Phi_j(p) - \Phi_j(q)] + w_k[\Phi_k(p) - \Phi_k(q)] = 0. \tag{8}$$

By assumption, $\Phi_j(p) - \Phi_j(q) = \Phi_k(q) - \Phi_k(p) \neq 0$. Thus, (8) implies that $w_j = w_k$. As this conclusion holds for any distinct $j, k \in N$, it follows that the welfare weights are all equal (and positive). Dividing both sides of the inequality in (7) by this common welfare weight, (7) simplifies to (6). □

Our assumption that the functions $\{\Phi_i\}$, $i \in N$, are the utility functions to be used for the purpose of social evaluation allows us to give a welfarist interpretation to the relative utilitarian order axiomatized in Theorem 2. It should be stressed that this axiology is only welfarist from the perspective of the social observer. It is him that determines what constitutes the normatively significant way of measuring the individual well-beings.

It could be argued that there is some independent basis for measuring strengths of preference and utility differences. For example, individual i could have an objectively determined cardinally-significant utility function U_i^* on $\Delta(X)$ that measures his ex ante well-being. That is, there is a fact of the matter about how to measure individual well-beings on an interval scale. Unless $U_i^* = \Phi_i$ for all i (which is implausible), our version of relative utilitarianism is not welfarist with respect to these objectively-given measures of well-being. However, such an interpretation of relative utilitarianism may not have normative appeal because it is sensitive to the differing capacities individuals have for achieving well-being. In our approach, these differences have no normative significance.[18]

In the single-profile approach to social choice theory, a social preference is determined for a single profile of individual preferences on some abstract set of alternatives A. The characterization theorems in this literature suppose that for any way that it is possible for the set of individuals to order three alternatives, there are three alternatives in A for which this pattern is realized.[19] This is a strong preference diversity assumption. In contrast, the single-profile characterization of relative utilitarianism in Theorem 2 only employs a mild preference diversity assumption, namely, Independent Prospects. The structure imposed on the set of alternatives and on the set of preferences by the expected utility model obviates the need for a stronger preference diversity assumption.

A notable feature of Theorem 2 is that the equality of the social weights is obtained in a single-profile setting, albeit one in which preferences are supplemented with information about strengths of preference. The standard way to obtain equal social weights is to use a multi-profile framework in which it is possible to permute the individuals' preferences or utility functions. By adding an anonymity axiom that requires the social preference to be invariant to such a permutation to any axiomatization of a weighted sum form of utilitarianism forces the weights all to be equal, as required by classical utilitarianism. An anonymity axiom is an interprofile condition. In criticizing Harsanyi (1955) for inappropriately using a symmetry argument in his single-profile setting in order to show that the weights in his Aggregation Theorem are all equal, Mongin and d'Aspremont (1998, p. 431, emphasis in the original) say that "it appears to be impossible to derive classical utilitarianism, i.e., equal weights utilitarianism, without

[18] Harsanyi (1955) offers three separate arguments in support of utilitarianism. In addition to the Social Aggregation Theorem considered here, Harsanyi defends utilitarianism using his Impartial Observer Theorem and using a theorem based on an assumption that there is a social ordering of vectors of individual utilities that is separable. When each individual has a normatively-significant measure of individual well-being that is not inferred from a von Neumann–Morgenstern representation, Blackorby et al. (1980) have shown that these three approaches to social evaluation may rank the social alternatives differently because they use different ordinal transforms of the normatively-significant utility functions.

[19] See Bossert and Weymark (2004, p. 1110).

imposing either [their anonymity axiom], or some variant *which must again be an interprofile condition*."[20] Our axiomatization of relative utilitarianism shows that this claim is too strong. By employing information about strengths of preference, not just preference rankings, we are able to obtain equal social weights without resorting to a multi-profile framework.

A related criticism of Harsanyi's utilitarian interpretation of his Aggregation Theorem that is discussed by Mongin and d'Aspremont (1998, p. 431) is that if this theorem is applied profile-by-profile in a multi-profile setting without imposing any interprofile conditions, then not only need the social weights not all be equal, they might also be profile-dependent. However, classical and weighted utilitarianism require that profile-independent weights be used to sum the individual utilities. Our axiomatization of relative utilitarianism applies to *any* profile of expected utility preferences that satisfies Independent Prospects. Consequently, if it is applied profile-by-profile in a multi-profile setting, the social weights are profile-independent (they are all equal) without the necessity of imposing any interprofile condition.

5 Alternative impartiality criteria

Impartiality is a moral imperative requiring that conflicting individual interests be socially resolved without favoring any one individual. Our impartiality axiom is closely related to other formalizations of this concept that have been proposed by Karni (1998, 2003) and Raschka (2023). In this section, we compare our impartiality axiom to theirs.

In Harsanyi's Impartial Observer Theorem (Harsanyi 1953), the social observer imagines being behind a veil of ignorance with an equal chance of being any individual once the veil is lifted. Harsanyi's Principle of Acceptance requires the social observer to agree with how i ranks two social alternative lotteries if he knows for certain that he will be person i once his identity is revealed. Karni and Weymark (1998) have argued that in order to invoke this principle, it is necessary to consider personal identity lotteries in which the probability of being any particular individual once the veil is lifted need not be the same for all individuals and to consider alternatives in which different individuals face different social alternative lotteries, what we call allocations. Formally, an *allocation* is a list of n social alternative lotteries in $\Delta(X)^n$, the ith of which is the one designated for person i. Individuals have preferences over their own lotteries in $\Delta(X)$.

Karni (1998) uses this analytical framework to define what may be described as being an *ordinal, or intrinsic, concept of impartiality*.[21] He makes the normative assumption that for all $\lambda \in [0, 1]$, the utility obtained by the lottery $\lambda\delta_{\hat{x}_i} + (1 - \lambda)\delta_{\check{x}_i}$ is the same for all $i \in N$. This is a claim about interpersonal comparisons of utility levels. Ordinal interpersonal comparisons of utility levels for arbitrary social alternative lotteries are facilitated by singling out one individual, say person 1, to

[20] See also Mongin (1994, p. 347).

[21] The axiomatization of relative utilitarianism in Karni (1998) requires that $|X|$ be non-finite but his impartiality axiom does not. We describe his axiom for finite X. Karni does not suppose that utilities are 0-1 normalized. He weights each individual's utility by the inverse of the range of his utility function before summing, which effectively makes his weighted utilitarian rule relative utilitarianism.

anchor them. For any $p \in \Delta(X)$, there is a unique $\lambda_p^1 \in [0,1]$ for which $p \sim_1 [\lambda_p^1 \delta_{\hat{x}_1} + (1-\lambda_p^1)\delta_{\check{x}_1}]$. Note that $p \succeq_1 q \leftrightarrow \lambda_p^1 \geq \lambda_q^1$. For any $i \in N$, let $\Psi_i(p)$ be any lottery in $\Delta(X)$ for which $\Psi_i(p) \sim_i [\lambda_p^1 \delta_{\hat{x}_i} + (1-\lambda_p^1)\delta_{\check{x}_i}]$. Because λ_p^1 is used for both 1 and i when mixing between their best and worst outcomes, i's utility with $\Psi_i(p)$ is the same as 1's is with p. Now, consider any $p, q \in \Delta(X)$ for which $p \succ_1 q$. Let a^1 and a^2 be two allocations for which (i) a^1 assigns $\Psi_j(p)$ to j and $\Psi_k(q)$ to k, (ii) a^2 assigns them $\Psi_j(q)$ and $\Psi_k(p)$, respectively, and (ii) a^1 and a^2 assign the same social alternative lotteries to everyone else. By construction, j is better off with a^1, k is better off with a^2, and everybody else is indifferent between a^1 and a^2. Furthermore, measured in terms of person 1's utilities, the utility gain for j if a^2 is replaced by a^1 is equal to the utility loss for k with the reverse change. This is true whatever utility function is used to represent $\succeq_1$, so no assumption is being made about strengths of preference. Karni argues that j and k should be treated impartially in such a comparison. This is accomplished by requiring the social observer to be indifferent between a^1 and a^2.

Karni (2003) is concerned with an *extrinsic concept of impartiality*. For the case in which the set of social alternative lotteries is $\Delta(X)$, he introduces an extrinsically defined equivalence relation $\approx$ on $\Delta(X)$ that determines in which two-person situations the conflicting interests of the two concerned individuals are of equal merit and, therefore, should be a matter of social indifference. For example, if (j, k, p, q) is a two-person situation and p and q are equivalent according to $\approx$, then p must be socially indifferent to q. In effect, $\approx$ is a partial ordering that indicates when two social alternative lotteries have equal social significance. The basis for choosing $\approx$ is not specified and so can be justified in different ways. One way to do so is to use our impartiality axiom. Because there is a single profile, if (j, k, p, q) is a two-person situation, there cannot be any other two-person situation in which p and q are the social alternative lotteries except for (k, j, q, p). If j's strength of preference for p over q is the same as k's for q over p, then Impartiality implies that $p \sim_0 q$ and $q \sim_0 p$. We can use this social indifference to define $\approx$ by setting $p \approx q \leftrightarrow p \sim_0 q$ when this is the case. It is easy to verify that $\approx$ so defined is an equivalence relation.[22]

In our model, an individual's strength of preference is defined using that person's ordinal preferences over $\Delta(X)$ and a difference in his well-being is identified with his strength of preference. In contrast, Raschka (2023) employs an extrinsic approach to well-being differences in a model in which the set of social alternatives is an arbitrary set X. He posits the existence of a binary relation $\succsim$ on the set $(N \times X)^2$ that provides a ranking of well-being differences. The statement $((i, x), (j, y)) \succsim ((k, z), (l, w))$ is interpreted as saying that the difference in the well-being of individual i when the social outcome is x and that of individual j when the social outcome is y is at least as large as that between individual k when the social outcome is z and individual l when the social outcome is w.

[22] Karni (2003) also develops a version of his extrinsic concept of impartiality for the framework employed in Karni (1998) in which the set of allocations is $\Delta(X)^n$ and individuals have preferences over $\Delta(X)$. In this case, the equivalence relation is over allocations and it is used to determine when a conflict of interests between two individuals are a matter of social indifference when there are only two concerned individuals.

It is natural to interpret $((i, x), (j, y)) \succsim ((i, x), (i, x))$ to mean that i is as well off with x as j is with y. Thus, $\succeq$ also allows comparisons levels of well-being levels. Hence, $\succeq$ induces a binary relation, $\succsim^*$, on $N \times X$ for which $(i, x) \succsim^* (j, y)$ means that the well-being of individual i when the social outcome is x is at least that of individual j when the social outcome is y. By restricting $\succsim^*$ to comparisons involving only individual i, we obtain a binary relation $\succeq_i$ on X that is the analogue of his individual preference in our model. The social preference $\succeq_0$ is also on X. Raschka's model is single-profile in the sense that it only considers one well-being difference relation and one social relation.

Raschka argues that in a situation (j, k, x, y) in which there are two concerned individuals with conflicting interests as measured by the preferences $\succeq_i$ (the analogue in his model of a two-person situation), if it is not the case that $x \sim_0 y$, then this is because either (i) the well-being differences of the two concerned individuals are different or (ii) one of them is worse off than the other in x or y. If level comparisons are precluded, it follows that if in (j, k, x, y) the well-being differences of j and k are of equal magnitude, then $x \sim_0 y$, which is Raschka's analogue of our impartiality axiom.[23] His impartiality axiom differs from ours because his well-being difference comparisons are based on the extrinsic relation $\succeq$, whereas ours are constructed intrinsically from individual preferences that satisfy the expected utility axioms.

6 Börgers and Choo's revealed preference approach

In this section, we describe the revealed preference approach Börgers and Choo (2017b) use to axiomatize relative utilitarianism and relate it to our own approach.

Consider the two-person situation (j, k, p, q) and suppose that $p \sim_0 q$. For this two-person situation, Börgers and Choo define the *marginal rate of substitution between i and j at p and q* as

$$\mathrm{MRS}_{jk}(p, q) = -\left[\frac{\Phi_j(p) - \Phi_j(q)}{\Phi_k(p) - \Phi_k(q)}\right]. \tag{9}$$

Using a 0-1 normalized von Neumann–Morgenstern utility function, the numerator on the right-hand side of (9) measures how much j's utility increases when q is replaced by p. Similarly, the denominator measures how much k's utility decreases with this change. Because the social observer and everybody except for j and k is indifferent between p and q, the social preference $\succeq_0$ can be interpreted as revealing that the social observer is willing to trade off the utilities of j and k at the rate $\mathrm{MRS}_{jk}(p, q)$ when q is replaced by a social alternative lottery p socially indifferent to it.[24]

The marginal rate of substitution in (9) is only defined for two-person situations (j, k, p, q) for which the social observer is indifferent between p and q. Börgers and Choo show that for any distinct pair of individuals, their definition applies to at least

[23] Raschka uses this condition as part of his axiomatization of classical utilitarianism.

[24] When Börgers and Choo introduce their definition of $\mathrm{MRS}_{jk}(p, q)$, they do not interpret $\Phi_j(p)$ as j's utility at p but, rather, as the probability of him obtaining his most preferred outcome in the social alternative lottery that is indifferent to p that only puts positive probability on j's best and worst outcomes, as in (5).

one pair of social alternative lotteries if Independent Prospects and Strong Pareto are satisfied. They further show that if $\succeq_0$ is represented by a utility function of the form $\sum_{i=1}^{n} w_i v_i$, where v_i is a von Neumann–Morgenstern utility representation of $\succeq_i$, then in any two-person situation (j, k, p, q) for which $p \sim_0 q$,

$$\text{MRS}_{jk}(p, q) = -\frac{w_k}{w_j}. \tag{10}$$

Hence, the social observer reveals that he is trading off the two concerned individuals' utilities using the ratio of their social weights in Harsanyi's aggregation equation (4).

Relative utilitarianism requires these weights to be equal when 0-1 normalized utility functions are used to represent the individual preferences. To obtain this outcome, Börgers and Choo extend their single-profile analysis to a multi-profile setting with a restricted domain of preference profiles and require that some interprofile conditions are satisfied.

With relative utilitarianism for 0-1 normalized individual utility functions, the weights in (10) are both 1. When this is the case, in Börgers and Choo's approach, the social observer's preference can be interpreted as revealing that the strengths of preference for p over q (as measured using Definition 2) of the concerned individuals are of equal magnitude but of opposite sign in any two-person situation (j, k, p, q) for which $p \sim_0 q$. In contrast, with our approach, the inference goes the other way. Our impartiality axiom implies that there is social indifference in any two-person situation (j, k, p, q) in which the strengths of preference for p over q of the concerned individuals are of equal magnitude but of opposite sign.

7 Concluding remarks

Our axiomatization of relative utilitarianism has been obtained using the same setting as Harsanyi (1955) by supplementing his axioms with an impartiality axiom that requires there to be social indifference between two social alternative lotteries if there are only two concerned individuals, they have conflicting interests of equal strength, and everybody else is indifferent. This axiom is based on a normative assessment of how to use the individual preferences to measure strength of preference. As such, it is an intrinsic conception of impartiality. Karni (1998, 2003) and Raschka (2023) also base their conceptions of impartiality on the assessment of the merits of conflicting interests of two individuals over the ranking of a pair of social alternatives conditional on all of the other individuals being indifferent. The distinctive feature of our impartiality axiom is the criterion used to adjudicate between conflicting interests. In contrast to the extrinsic criteria employed by Karni (2003) and Raschka (2023), ours is is intrinsic. In contrast to the intrinsic criterion used by Karni (1998), ours concerns strengths of preference, not utility levels.

While we have offered a novel axiomatization of relative utilitarianism, we have not claimed that this axiomatization provides a compelling argument for employing relative utilitarianism to make collective decisions. All procedures for using information about the preferences or well-beings of individuals to determine a social ranking of the alternatives have their drawbacks—relative utilitarianism is no exception. As is the case with the Borda rule, with relative utilitarianism, the social ranking may be sensitive to how many alternatives there are. For example, suppose that $\tilde{X} = X \cup \{\tilde{x}\}$. Consider any lotteries $p, q \in X$ and $\tilde{p}, \tilde{q} \in \tilde{X}$ for which, for all $x \in X$, $\tilde{p}(x) = p(x)$ and $\tilde{q}(x) = q(x)$ and, hence, for which $\tilde{p}(\tilde{x}) = \tilde{q}(\tilde{x}) = 0$. Further suppose that (i) $\tilde{x}$ is uniquely best on $\tilde{X}$ for person 1, (ii) for any other individual, the best and worst alternatives on $\tilde{X}$ are the same as on X, and (iii) for all $i \in N$, i's ranking of any two lotteries in $\Delta(\tilde{X})$ for which there is no probability of obtaining $\tilde{x}$ is the same as his ranking of the corresponding lotteries in $\Delta(X)$. Except for person 1, the strength of preference for p over q is the same as that for $\tilde{p}$ over $\tilde{q}$. However, this is not the case for person 1 because his strength of preference is recalibrated when $\tilde{x}$ is added to X. Depending on the magnitude of his preference strength change, the social ranking of p and q could differ from that of $\tilde{p}$ and $\tilde{q}$. Such a social preference reversal is arguably an unsatisfactory feature of relative utilitarianism.

Acknowledgements We are grateful to Jean Baccelli, Jake Nebel, François Maniquet, our referees, and the participants in the Axiology 2023 conference held in Montreal for their comments.

References

Alt F (1936) Über die messbarkeit des nutzens. Zeitschrift für Nationalökonomie 7:161–169 (**Translated as Alt (1971)**)

Alt F (1971) On the measurement of utility. In: Chipman JS, Hurwicz L, Richter MK, Sonnenschein HF (eds) Preferences, utility, and demand: a Minnesota symposium. Harcourt Brace Jovanovich, New York, pp 424–431

Argenziano R, Gilboa I (2019) Perception-theoretic foundations of weighted utilitarianism. Econ J 129:1511–1528

Baccelli J (2024) Ordinal utility differences. Soc Choice Welf 62:275–287

Baccelli J, Mongin P (2016) Choice-based cardinal utility: a tribute to Patrick Suppes. J Econ Methodol 23:268–288

Binmore K (2009) Rational decisions. Princeton University Press, Princeton

Blackorby C, Donaldson D, Weymark JA (1980) On John Harsanyi's defences of utilitarianism. Discussion Paper No. 8013, Center for Operations Research and Econometrics, Université catholique de Louvain

Börgers T, Choo Y-M (2017a) A counterexample to Dhillon (1998). Soc Choice Welf 48:837–843

Börgers T, Choo Y-M (2017b) Revealed relative utilitarianism. Unpublished manuscript, Department of Economics, University of Michigan

Bossert W, Weymark JA (2004) Utility in social choice. In: Barberà S, Hammond PJ, Seidl C (eds) Handbook of utility theory. Volume 2: extensions. Kluwer Academic Publishers, Boston, pp 1099–1177

Brandl F (2021) Belief-averaging and relative utilitarianism. J Econ Theory 198:105368

Dhillon A (1998) Extended Pareto rules and relative utilitarianism. Soc Choice Welf 15:521–542

Dhillon A, Mertens J-F (1999) Relative utilitarianism. Econometrica 67:471–498

Edgeworth FY (1881) Mathematical psychics: an essay on the application of mathematics to the moral sciences. C. Kegan-Paul, London

Ellingsen T (1994) Cardinal utility: a history of hedonimetry. In: Allais M, Hagen O (eds) Cardinalism: a fundamental approach. Kluwer Academic Publishers, Dordrecht, pp 105–165

Harsanyi JC (1953) Cardinal utility in welfare economics and in the theory of risk-taking. J Polit Econ 61:434–435

Harsanyi JC (1955) Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. J Polit Econ 63:309–321

Karni E (1998) Impartiality: definition and representation. Econometrica 66:1405–1415

Karni E (2003) Impartiality and interpersonal comparisons of variations in well-being. Soc Choice Welf 21:95–111

Karni E, Weymark JA (1998) An informationally parsimonious impartial observer theorem. Soc Choice Welf 15:321–332

Luce RD, Raiffa H (1957) Games and decisions: introduction and critical survey. Wiley, New York

Maniquet F (2016) Social ordering functions. In: Adler MD, Fleurbaey M (eds) The Oxford handbook of well-being and public policy. Oxford University Press, New York, pp 227–245

Mongin P (1994) Harsanyi's aggregation theorem: multi-profile version and unsettled questions. Soc Choice Welf 11:311–354

Mongin P, d'Aspremont C (1998) Utility theory and ethics. In: Barberà S, Hammond PJ, Seidl C (eds) Handbook of utility theory. Volume 1: principles. Kluwer Academic Publishers, Boston, pp 371–481

Montesano A (1985) The ordinal utility under uncertainty and the measure of risk aversion in terms of preferences. Theor Decis 18:73–85

Moscati I (2019) Measuring utility: from the marginal revolution to behavioral economics. Oxford University Press, New York

Nebel JM (2021) Utils and shmutils. Ethics 131:571–599

Nebel JM (2022) Aggregation without interpersonal comparisons of well-being. Philos Phenomenol Res 105:18–41

Ng Y-K (1975) Bentham or Bergson? Finite sensibility, utility functions and social welfare functions. Rev Econ Stud 42:545–569

Raschka R (2023) A single relation theory of welfarist social evaluation. Unpublished manuscript, Department of Economics, University of Hamburg

Robbins L (1938) Interpersonal comparisons of utility: A comment. Econ J 48:635–641

Savage LJ (1954) The foundations of statistics. Wiley, New York

Segal U (2000) Let's agree that all dictatorships are equally bad. J Polit Econ 108:569–589

Sen A (1976) Welfare inequalities and Rawlsian axiomatics. Theor Decis 7:243–262

Sprumont Y (2013) On relative egalitarianism. Soc Choice Welf 40:1015–1032

von Neumann J, Morgenstern O (1944) Theory of games and economic behavior. Princeton University Press, Princeton

Weymark JA (1991) A reconsideration of the Harsanyi-Sen debate on utilitarianism. In: Elster J, Roemer JE (eds) Interpersonal comparisons of well-being. Cambridge University Press, Cambridge, pp 255–320

Weymark JA (2005) Measurement theory and the foundations of utilitarianism. Soc Choice Welf 25:527–555