Arrovian Aggregation of Convex Preferences and Pairwise Utilitarianism

Florian Brandl Felix Brandt
Technische Universität München
{brandlfl,brandtf}@in.tum.de

We consider social welfare functions that satisfy Arrow's classic axioms of *independence of irrelevant alternatives* and *Pareto optimality* when the outcome space is the convex hull over some finite set of alternatives. Individual and collective preferences are assumed to be continuous and convex, which guarantees the existence of maximal elements and the consistency of choice functions that return these elements, even without insisting on transitivity. We provide characterizations of both the domains of preferences and the social welfare functions that allow for Arrovian aggregation. The domains allow for arbitrary preferences over alternatives, which in turn completely determine an agent's preferences over all outcomes. The only Arrovian social welfare functions on these domains constitute an interesting combination of utilitarianism and pairwiseness. When also assuming anonymity, Arrow's impossibility turns into a complete characterization of a unique social welfare function, which can be readily applied in settings that allow for lotteries or divisible resources such as time or money.

1. Introduction

A central concept in welfare economics are social welfare functions (SWFs) in the tradition of Arrow, i.e., functions that map a collection of individual preference relations over some set of alternatives to a social preference relation over the alternatives. Arrow's seminal theorem states that every SWF that satisfies Pareto optimality and independence of irrelevant alternatives is dictatorial (Arrow, 1951). This sweeping impossibility significantly strengthened an earlier observation by Condorcet (1785) and sent shockwaves throughout economic theory as well as political philosophy (see, e.g., Maskin and Sen, 2014; Sen, 2017). A large body of subsequent work has studied whether more positive results can be obtained by modifying implicit assumptions on the domain of admissible preferences, both individually and collectively. Two main approaches can be distinguished.

The first approach, pioneered by Sen (1969), has been to *weaken the assumption of transitivity of collective preferences* to quasi-transitivity, acyclicity, path independence, or similar conditions. Although this does allow for non-dictatorial aggregation functions that meet Arrow's criteria, these functions turned out to be highly objectionable, usually on grounds of involving weak kinds of dictatorships or violating other conditions deemed to be indispensable for reasonable preference aggregation (for an overview of the extensive literature, see Kelly, 1978; Sen, 1977, 1986; Schwartz, 1986; Campbell and Kelly, 2002). Particularly noteworthy are results about acyclic collective preference relations (e.g., Mas-Colell and Sonnenschein, 1972; Brown, 1975; Blau and Deb, 1977; Blair and Pollak, 1982; Blair and Pollack, 1983; Banks, 1995) because acyclicity is necessary and sufficient for the existence of maximal elements when there is a finite number of alternatives. Sen (1995) concludes that "the arbitrariness of power of which Arrow's case of dictatorship is an extreme example, lingers in one form or another even when transitivity is dropped, so long as *some* regularity is demanded (such as the absence of cycles)."

Another stream of research has analyzed the implications of *imposing additional structure on the individual preferences*. This has resulted in a number of positive results for restricted domains, such as dichotomous or single-peaked preferences, which allow for attractive SWFs (e.g., Black, 1948; Arrow, 1951; Inada, 1969; Sen and Pattanaik, 1969; Ehlers and Storcken, 2008). Many domains of economic interest are concerned with infinite sets of outcomes, which satisfy structural restrictions such as compactness and convexity. Preferences over these outcomes are typically assumed to satisfy some form of continuity and convexity, which roughly imply that preferences are robust with respect to minimal changes in outcomes and with respect to convex combinations of outcomes. Various results have shown that Arrow's impossibility remains intact under these assumptions (e.g., Kalai et al., 1979; Border, 1983; Bordes and Le Breton, 1989, 1990a,b; Campbell, 1989; Redekop, 1995). Le Breton and Weymark (2011) provide an overview and conclude that "economic domain restrictions do not provide a satisfactory way of avoiding Arrovian social choice impossibilities, except when the set of alternatives is one-dimensional and preferences are single-peaked."

The point of departure for the present approach is the observation that all these impossibilities involve some form of transitivity (e.g., acyclicity), even though no such assumption is necessary to guarantee the existence of maximal elements in domains of continuous and convex preferences. Sonnenschein (1971) has shown that all continuous and convex preference relations admit a maximal element in every non-empty, compact, and convex set of outcomes. Moreover, returning maximal elements under the given conditions satisfies standard properties of choice consistency introduced by Sen (1969, 1971). Continuous and convex preference relations can thus be interpreted as rationalizing relations for the choice behavior of rational agents. Consequently, there is little justification for demanding transitivity, which has come under independent attack in normative and descriptive decision theory (see, e.g., May, 1954; Tversky, 1969; Fishburn, 1970; Bar-Hillel and Margalit, 1988;

Fishburn, 1991; Anand, 1993, 2009).[1] For example, the *preference reversal phenomenon*, first described by Grether and Plott (1979), shows systematic experimental failures of transitivity where one lottery is preferred to another but the certainty equivalent of the latter is preferred to that of the former. As Anand (2009) writes, "once considered a cornerstone of rational choice theory, the status of transitivity has been dramatically reevaluated by economists and philosophers in recent years."

Summary of Results

We show that Arrow's theorem ceases to hold on convex domains when dispensing with transitivity, and, moreover, Arrow's axioms, along with some weak technical assumptions, narrow down the choice of a suitable SWF to an interesting form of *pairwise utilitarianism*. The SWF is utilitarian because collective preferences are obtained by adding the canonical skew-symmetric bilinear (SSB) utility functions representing the voters' ordinal preferences. SSB utility theory is more general than traditional linear utility theory since SSB functions assign a utility value to each *pair* of alternatives. Hence, the SWF merely takes into account the numbers of agents who prefer one alternative to another and also satisfies Condorcet's pairwise majority criterion.

More precisely, we consider a convex set of outcomes consisting of all probability measures on some abstract set of alternatives, which we refer to as pure outcomes. These outcome sets arise when allocating a divisible resource (such as probability, time, or money) to alternatives. A natural universal example is the standard unstructured social choice setting that also allows for lotteries between alternatives. Individual and collective preference relations over these outcomes are assumed to satisfy continuity, convexity, and symmetry. We then show that there is a unique inclusion-maximal Cartesian domain of preference profiles that allows for anonymous Arrovian aggregation while satisfying minimal richness conditions. This domain allows for arbitrary preferences over pure outcomes, which in turn completely determine an agent's preferences over all remaining outcomes. When interpreting outcomes as lotteries, this preference extension has a particularly simple and intuitive explanation: *an agent prefers one lottery to another if and only if the former is more likely to return a more preferred alternative*. Incidentally, this preference extension, which constitutes a central special case of SSB utility functions as introduced by Fishburn (1982), is supported by recent experimental evidence.

We then prove that the only Arrovian SWFs are affine utilitarian with respect to the underlying SSB utility functions. As a consequence, there is a unique anonymous Arrovian

[1] Another frequently cited reason to justify transitivity is the *money pump*, where an agent with cyclic preferences over three outcomes is deceived into paying unlimited amounts of money in an infinite series of cyclical exchanges. As Fishburn (1991) notes, however, the money pump "applies transitive thinking [in the form of money] to an intransitive world". Another issue with the money pump in our framework is that it cleverly avoids convexity of the feasible set by splitting it up into three subsets whose union is not convex. If the agent were confronted with a choice from the convex hull of the three original outcomes, he could simply pick his (unique) most-preferred mixed outcome and would not be tempted to exchange it when offered any other outcome in the future.

SWF, which compares outcomes by the sign of the bilinear form given by the pairwise majority margins. The resulting collective preference relation over *pure* outcomes coincides with majority rule and the corresponding choice function is therefore consistent with Condorcet's principle of selecting a pure outcome that is majority-preferred to every other pure outcome whenever this is possible.[2] This relation is naturally extended to mixed outcomes such that, by the Minimax Theorem, every compact and convex set of outcomes admits a collectively most preferred outcome.[3]

2. Related Work

A special case of our setting, which has been well studied, deals with individual preferences over lotteries that satisfy the *von Neumann-Morgenstern (vNM) axioms*, i.e., preferences that can be represented by assigning cardinal utilities to alternatives such that lotteries are compared based on the expected utility they produce. Samuelson (1967) conjectured that Arrow's impossibility still holds under these assumptions and Kalai and Schmeidler (1977b) showed that this is indeed the case when there are at least four alternatives. Hylland (1980) later pointed out that a continuity assumption made by Kalai and Schmeidler is not required. There are other versions of Arrow's impossibility for vNM preferences, which differ in modeling assumptions and whether SWFs aggregate cardinal utilities or the preference relations represented by these utilities. A detailed comparison of these results is given in Appendix A.

Our results apply to Arrovian aggregation of preferences over lotteries under much loosened assumptions about preferences over lotteries. In particular, the axioms we presume entail that preferences over lotteries can be represented by *skew-symmetric bilinear (SSB) utility functions*, which assign a utility value to each pair of lotteries. One lottery is preferred to another lottery if the SSB utility for this pair is positive. SSB utility theory is a generalization of linear expected utility theory due to von Neumann and Morgenstern (1947), which does not require the controversial independence axiom and transitivity (see, e.g., Fishburn, 1982, 1984b, 1988). Independence prescribes that a lottery p is preferred to lottery q if and only if a coin toss between p and a third lottery r is preferred to a coin toss between q and r (with the same coin used in both cases). There is experimental evidence that independence is systematically violated by human decision makers. The Allais Paradox is perhaps the most famous example (Allais, 1953). Detailed reviews of such violations, including those reported by Kahnemann and Tversky (1979), have been

[2] It is therefore in line with Dasgupta and Maskin (2008) who, also based on Arrow's axioms, have forcefully argued in favor of majority rule in domains where Condorcet winners are guaranteed to exist. Our arguments extend to unrestricted preferences over pure outcomes.

[3] Since our domain allows for *arbitrary* preference relations over pure outcomes, it encapsulates well-known domains that allow for Arrovian aggregation in non-convex settings (such as dichotomous and single-peaked preferences). In these subdomains, the collective preference relation over pure outcomes happens to be transitive and therefore guarantees maximal pure outcomes.

provided by Machina (1983, 1989) and McClennen (1988).[4]

Our characterization of Arrovian SWFs is related to Harsanyi's *Social Aggregation Theorem* (Harsanyi, 1955), which shows that, for von Neumann-Morgenstern preferences over lotteries, affine utilitarianism already follows from Pareto indifference (see Fleurbaey et al., 2008, for an excellent exposition and various extensions of this theorem). Harsanyi's theorem is a statement about *Bergson-Samuelson social welfare functions*, i.e., a single preference profile is considered in isolation. As a consequence, the weights assigned to the agents' utility functions may depend on their preferences. This can be prevented by adding axioms that connect the collective preferences across different profiles. The SWF that derives the collective preferences by adding up utility representations normalized to the unit interval is known as *relative utilitarianism* (Dhillon, 1998; Dhillon and Mertens, 1999; Börgers and Choo, 2015, 2017). It was characterized by Dhillon and Mertens (1999) using essentially *independence of redundant alternatives* (a weakening of independence of irrelevant alternatives) and monotonicity (a weakening of a Pareto-type axiom). As shown by Fishburn and Gehrlein (1987) and further explored by Turunen-Red and Weymark (1999), aggregating SSB utility functions is fundamentally different from aggregating von Neumann-Morgenstern utility functions in that Harsanyi's Pareto indifference axiom (and strengthenings thereof) do not imply affine utilitarianism. As we show in this paper, this can be rectified by considering Arrow's multi-profile framework and assuming independence of irrelevant alternatives. Mongin (1994, Proposition 1) gives a similar characterization of affine utilitarianism for social welfare functionals (SWFLs), which operate on profiles of cardinal utility functions. The framework of SWFLs (without any invariance property on utility functions) involves interpersonal comparisons of utilities, which significantly weakens the force of IIA (see also Appendix A).

The probabilistic voting rule that returns the maximal elements of the unique anonymous Arrovian SWF is known as *maximal lotteries* (Kreweras, 1965; Fishburn, 1984a) and was recently axiomatized using two consistency conditions (Brandl et al., 2016). Independently, maximal lotteries have also been studied in the context of randomized matching and assignment (see, e.g., Kavitha et al., 2011; Aziz et al., 2013).

When the set of outcomes cannot be assumed to be convex (for example, because it is finite), a common approach to address the intransitivity of collective preferences is to define alternative notions of maximality, rationalizability, or welfare, leading to concepts such as *transitive closure maximality* or the *uncovered set* (see, e.g., Laslier, 1997; Brandt and Harrenstein, 2011; Brandt et al., 2017; Nishimura, 2017). Interestingly, the support of maximal lotteries, which is known as the *bipartisan set* or the *essential set* (Laffond et al., 1993; Laslier, 2000), also appears in this literature, even though this approach is fundamentally different from the one pursued in this paper.

[4] Fishburn and Wakker (1995) give an interesting historical perspective on the independence axiom.

3. Preliminaries

Let U be a non-empty, finite universal set of alternatives. By Δ we denote the set of all probability measures on U. For $X \subseteq U$, let Δ_X be the set of probability measures in Δ with support in X, i.e., $\Delta_X = \{p \in \Delta \colon p(X) = 1\}$. We will refer to elements of Δ as *outcomes* and one-point measures in Δ as *pure outcomes*. Furthermore, let $\succ$ be an asymmetric binary relation over Δ, which is interpreted as the *preference relation* of an agent. Given two outcomes $p, q \in \Delta$, we write $p \sim q$ when neither $p \succ q$ nor $q \succ p$, and $p \succsim q$ if $p \succ q$ or $p \sim q$. For $p \in \Delta$, let $U(p) = \{q \in \Delta \colon q \succ p\}$ and $L(p) = \{q \in \Delta \colon p \succ q\}$ be the *strict upper* and *strict lower contour set* of p with respect to $\succ$; $I(p) = \{q \in \Delta \colon p \sim q\}$ denotes the *indifference set* of p. For $X \subseteq \Delta$, $\succ|_X = \{(p, q) \in \succ \colon p, q \in X\}$ is the preference relation $\succ$ restricted to outcomes in X.

We will consider preference relations that are continuous, i.e., small changes in outcomes do not result in a reversal of preference. One way to model this is the Archimedean axiom, which requires that, for any given outcome q, the convex hull of a more preferred outcome and a less preferred outcome also contains an equally preferred outcome. A preference relation $\succ$ is continuous if, for all $p, q, r \in \Delta$,

$$p \succ q \succ r \text{ implies } \lambda p + (1-\lambda) r \sim q \text{ for some } \lambda \in (0,1). \tag{Continuity}$$

Another standard assumption is that preferences are convex. We will use convexity as defined by Fishburn (1982). A preference relation $\succ$ is convex if, for all $p, q, r \in \Delta$ and $\lambda \in (0,1)$,

$$\begin{aligned} &p \succ q \text{ and } p \succsim r \text{ imply } p \succ \lambda q + (1-\lambda) r,\\ &q \succ p \text{ and } r \succsim p \text{ imply } \lambda q + (1-\lambda) r \succ p, \text{ and}\\ &p \sim q \text{ and } p \sim r \text{ imply } p \sim \lambda q + (1-\lambda) r. \end{aligned} \tag{Convexity}$$

Equivalently, one could require that the indifference set for an outcome p is the intersection of Δ with a hyperplane through p; the upper and lower contour sets are the intersection of Δ with the corresponding half spaces. Note that convexity implies that upper contour sets, lower contour sets, and indifference sets are convex. Moreover, upper contour and lower contour sets are open and indifference sets are closed.

The existence of maximal elements is usually quoted as the main reason for insisting on transitivity of preference relations. It was shown by Sonnenschein (1971) that continuity and convexity are already sufficient for the existence of maximal elements, even when preferences are intransitive (see also Bergstrom, 1992; Llinares, 1998).[5] For a preference relation $\succ$ and a subset of outcomes $X \subseteq \Delta$, let $\max_\succ(X) = \{x \in X \colon x \succsim y \text{ for all } y \in X\}$.

Proposition 1. *(Sonnenschein, 1971) If $\succ$ is a continuous and convex preference relation, then $\max_\succ(X) \neq \emptyset$ for every non-empty, compact, and convex set $X \subseteq \Delta$.*

[5] Sonnenschein only required that upper contour sets are convex and that lower contour sets are open. Also his notion of continuity is weaker than ours when the set of alternatives is finite.

Sen (1969, 1971) has shown that two intuitive choice consistency conditions are equivalent to choosing maximal elements according to an acyclic relation. These conditions are known as *Sen's* α (or *contraction*) and *Sen's* γ (or *expansion*). Contraction requires that if an outcome is chosen from some set, then it is also chosen from any subset that it is contained in. This condition is satisfied by $\max_\succ$ without imposing any restrictions on $\succ$. Expansion prescribes that an outcome that is chosen from two sets X and Y, should also be chosen from their union $X \cup Y$. Since we are only interested in choosing from convex sets, we strengthen this condition by taking the *convex hull* $\text{conv}(X \cup Y)$ in the consequence. $\max_\succ$ satisfies this condition whenever $\succ$ is convex. To see this, consider $X, Y \subseteq \Delta$ and assume that $p \in \max_\succ X \cap \max_\succ Y$. Then, $p \succsim q$ for all $q \in X \cup Y$ and since $\succ$ satisfies convexity, we have $p \succsim q$ for all $q \in \text{conv}(X \cup Y)$. Thus, $p \in \max_\succ(\text{conv}(X \cup Y))$. Sen's proof can even be adapted to show that *every* choice function satisfying contraction and expansion is of the form $\max_\succ$ for some $\succ$ with convex weak upper and lower contour sets.[6]

Convexity of preferences implies that indifference curves are straight lines. The symmetry axiom introduced by Fishburn (1982) prescribes that the indifference curves for every triple of outcomes are parallel or intersect in one point, which may be outside of their convex hull. For all $p, q, r \in \Delta$ and $\lambda \in (0, 1)$,

$$\begin{gathered}\text{if } p \succ q \succ r,\ p \succ r, \text{ and } q \sim {}^1\!/_2\, p + {}^1\!/_2\, r, \text{ then} \\ [\lambda p + (1-\lambda) r \sim {}^1\!/_2\, p + {}^1\!/_2\, q \text{ if and only if } \lambda r + (1-\lambda) p \sim {}^1\!/_2\, r + {}^1\!/_2\, q].\end{gathered} \tag{Symmetry}$$

Fishburn (1984b) justifies this axiom by stating that "the degree to which p is preferred to q is equal in absolute magnitude but opposite in sign to the degree to which q is preferred p." He continues by writing that he is "a bit uncertain as to whether this should be regarded more as a convention than a testable hypothesis – much like the asymmetry axiom [...], which can almost be thought of as a definitional characteristic of strict preference."

By $\mathcal{R}$ we denote the set of all continuous, convex, and symmetric preference relations. Despite the richness of $\mathcal{R}$, preference relations therein admit a particularly nice representation. It was shown by Fishburn (1982) that if $\succ \in \mathcal{R}$, then there is a skew-symmetric and bilinear (SSB) utility function $\phi\colon \Delta \times \Delta \to \mathbb{R}$ such that, for all $p, q \in \Delta$,[7]

$$p \succ q \text{ if and only if } \phi(p, q) > 0.$$

Moreover, ϕ is unique up to scalar multiplication. We therefore write $\phi \equiv \hat{\phi}$ if and only if there is some $\alpha > 0$ such that $\phi = \alpha \cdot \hat{\phi}$, i.e., if ϕ and $\hat{\phi}$ represent the same preferences. We will also write $\succ \equiv \phi$ if $\succ$ is represented by the SSB function ϕ. Let Φ denote the set of all SSB functions on $\Delta \times \Delta$. Since all outcomes have finite support, $\phi(p, q)$ can be

[6] There are also stronger versions of expansion, which, together with contraction, are equivalent to the *weak axiom of revealed preference* or *Arrow's choice axiom* (Samuelson, 1938; Arrow, 1959). These conditions imply rationalizability via a *transitive* relation and are therefore not generally satisfied when choosing maximal elements of convex relations.

[7] A function ϕ is skew-symmetric if $\phi(p, q) = -\phi(q, p)$ for all $p, q \in \Delta$. ϕ is bilinear if it is linear in both arguments.

written as a convex combination of the values of ϕ for pure outcomes (Fishburn, 1984b). For this purpose, we identify every alternative $a \in U$ with the pure outcome that assigns probability 1 to a. Then, for all $p, q \in \Delta$,

$$\phi(p,q) = \sum_{a,b \in U} p(a)q(b)\phi(a,b).$$

We will often represent SSB functions restricted to Δ_X for $X \subseteq U$ as skew-symmetric matrices in $\mathbb{R}^{X \times X}$.

When requiring transitivity on top of continuity, convexity, and symmetry, the four axioms characterize *weighted linear (WL)* utility functions as introduced by Chew (1983).[8] When additionally requiring independence, then ϕ is separable, i.e., $\phi(p,q) = u(p) - u(q)$, where u is a linear von Neumann-Morgenstern utility function representing $\succ$. For independently distributed outcomes (as considered in this paper), SSB utility theory coincides with regret theory as introduced by Loomes and Sugden (1982) (see also Loomes and Sugden, 1987; Blavatskyy, 2006).

Through the representation of $\succ \in \mathcal{R}$ as a skew-symmetric matrix, it becomes apparent that the Minimax Theorem implies the existence of maximal elements of $\succ$ on Δ_X. This was noted by Fishburn (1984b, Theorem 4) and already follows from Proposition 1. The matrix representation also yields a simple way to see that the indifference curves for any triple of outcomes intersect in one point within their affine hull. This point is given by the null space of the corresponding matrix. Note that the null space of a skew-symmetric matrix of odd size cannot be empty. If the triple is transitively ordered, then this point is outside of their convex hull, otherwise it is inside.

4. Social Welfare Functions

In the remainder of this paper we deal with the problem of aggregating the preferences of multiple agents into a collective preference relation. The set of agents is $N = \{1, \dots, n\}$ for some $n \geq 2$. The preference relations of agents belong to some *domain* $\mathcal{D} \subseteq \mathcal{R}$. A function $R \in \mathcal{D}^N$ from the set of agents to the domain is a preference profile. We will write preference profiles as tuples with indices in N. A *social welfare function (SWF)* $f \colon \mathcal{D}^N \to \mathcal{R}$ maps a preference profile to a collective preference relation.[9]

Arrow (1951) considered the unrestricted domain of all complete and transitive preferences over some set of alternatives and showed that the only SWFs that satisfy two desirable properties, Pareto optimality and independence of irrelevant alternatives, are dictatorial

[8] A WL function is characterized by a linear utility function and a linear and non-negative weight function. The utility of an outcome is the utility derived by the linear utility function weighted according to the weight function. Thus, WL functions are more general than linear utility functions, as every linear utility function is equivalent to a WL function with constant weight function. See also Fishburn (1983).

[9] The domains we consider are therefore Cartesian and identical for all players. This, for example, rules out the domain of single-peaked preferences.

functions. Pareto optimality prescribes that a unanimous preference of one outcome over another in the individual preferences should be reflected in the collective preference. An SWF f satisfies *Pareto optimality* if, for all $p, q \in \Delta$, $R \in \mathcal{D}^N$, and $f(R) = \succ$,

$$\begin{aligned} &p \succsim_i q \text{ for all } i \in N \text{ implies } p \succsim q, \text{ and} \\ &\text{if additionally } p \succ_i q \text{ for some } i \in N \text{ then } p \succ q. \end{aligned} \tag{Pareto optimality}$$

The indifference part of Pareto optimality, which merely requires that $p \sim_i q$ for all $i \in N$ implies $p \sim q$, is usually referred to as *Pareto indifference*.

Independence of irrelevant alternatives demands that collective preferences over some feasible set of outcomes should only depend on the individual preferences over this set (and not on the preferences over outcomes outside this set). In our framework, we will assume that feasible sets are based on the availability of alternatives and are therefore of the form Δ_X for $X \subseteq U$ (see also Kalai and Schmeidler, 1977b). Formally, we say that an SWF f satisfies *independence of irrelevant alternatives (IIA)* if, for all $R, \hat{R} \in \mathcal{D}^N$ and $X \subseteq U$,

$$R|_{\Delta_X} = \hat{R}|_{\Delta_X} \text{ implies } f(R)|_{\Delta_X} = f(\hat{R})|_{\Delta_X}. \tag{IIA}$$

Any SWF that satisfies Pareto optimality and IIA will be called an *Arrovian* SWF. Arrow has shown that, when no structure is imposed on preference relations and feasible sets, every Arrovian SWF is dictatorial, i.e., there is $i \in N$ such that for all $p, q \in \Delta$, $R \in \mathcal{D}^N$, and $f(R) = \succ$, $p \succ_i q$ implies $p \succ q$. Dictatorships are examples of SWFs that are extremely biased towards one agent. In many applications, *any* differentiation between agents is unacceptable and all agents should be treated equally. This property is known as anonymity. We denote by Π_N the set of all permutations on N. For $\pi \in \Pi_N$ and a preference profile $R \in \mathcal{D}^N$, $R^\pi = R \circ \pi$ is the preference profile where agents are renamed according to π. Then, an SWF f satisfies *anonymity* if, for all $R \in \mathcal{D}^N$ and $\pi \in \Pi_N$,

$$f(R) = f(R^\pi). \tag{Anonymity}$$

Anonymity is obviously a stronger requirement than non-dictatorship.

In order to prove our characterization, we need to assume that the domain $\mathcal{D} \subseteq \mathcal{R}$ satisfies certain richness conditions. First, we require that it is neutral in the sense that it is not biased towards certain alternatives. For $\pi \in \Pi_U$ and $p \in \Delta$, let $p^\pi \in \Delta$ such that $p^\pi(\pi(a)) = p(a)$ for all $a \in U$. Then, for $\succ \in \mathcal{R}$, we define $\succ^\pi$ such that $p^\pi \succ^\pi q^\pi$ if and only if $p \succ q$ for all $p, q \in \Delta$. It is assumed that $\succ \in \mathcal{D}$ if and only if $\succ^\pi \in \mathcal{D}$ for all $\pi \in \Pi_U$ and $\succ \in \mathcal{D}$. Second, it should also be possible for agents to declare completely opposed preferences. For $\succ \in \mathcal{D}$, $\succ^{-1}$ is the inverse of $\succ$, i.e., $p \succ^{-1} q$ if and only if $q \succ p$ for all $p, q \in \Delta$. Then $\succ \in \mathcal{D}$ implies $\succ^{-1} \in \mathcal{D}$ for all $\succ \in \mathcal{R}$. Note that this condition is not implied by the previous neutrality condition because it allows the inversion of preferences over all outcomes, not only over pure outcomes. Finally, we demand that for every transitive and asymmetric relation on *pure* outcomes, $\mathcal{D}$ contains at least one extension of this relation to all outcomes in Δ. The last assumption can be slightly weakened without affecting the correctness of our proofs (see Remark 5).

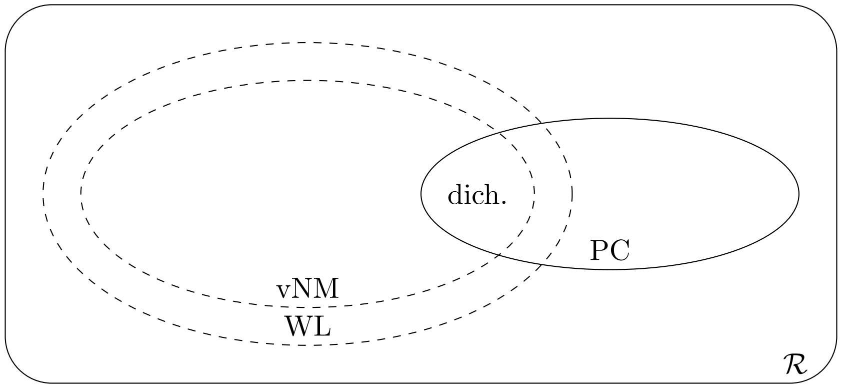

Figure 1: Venn diagram showing the inclusion relationships between preference domains. The intersection of the domain of vNM preferences and the domain of PC (pairwise comparison) preferences contains exactly the set of dichotomous vNM preference relations. The intersection of WL (weighted linear utility) preferences and PC preferences contains exactly the PC preferences based on trichotomous weak orders (see Figure 2 for an example). An example of PC preferences not contained in the set of WL preferences is given in Figure 3. Theorem 1 shows that the domain of PC preferences is the unique inclusion-maximal domain for which anonymous Arrovian aggregation is possible within $\mathcal{R}$. This, for example, implies impossibilities for WL preferences and vNM preferences.

5. Characterization of the Domain

Non-dictatorial Arrovian aggregation on the full domain $\mathcal{R}$ is impossible because it is already impossible in the subdomain of vNM preferences (see Appendix A). On the other hand, interesting possibilities emerge in restricted domains such as in that of dichotomous vNM preferences where each agent can only assign two different utility values. In this domain, every SWF based on affine welfare (with positive weights) satisfies IIA and Pareto optimality. The only anonymous Arrovian SWF on this domain corresponds to *approval voting* and ranks alternatives by the number of approvals they receive from the agents. This ranking is identical to majority rule, which happens to be transitive for dichotomous preferences.

In this section, we characterize the unique inclusion-maximal domain $\mathcal{D} \subseteq \mathcal{R}$ for which anonymous Arrovian SWFs exist. We say that $\phi \in \Phi$ is based on *pairwise comparisons* if $\phi(a,b) \in \{-1,0,1\}$ for all $a,b \in U$ and denote the set of SSB functions that are based on pairwise comparisons by $\Phi^{PC} \subset \Phi$ and the corresponding set of preference relations by $\mathcal{D}^{PC} = \{\succ \in \mathcal{R} \colon \succ \equiv \phi \text{ for some } \phi \in \Phi^{PC}\}$.

Theorem 1. *Let f be an anonymous Arrovian SWF on some domain $\mathcal{D}$ with $|U| \geq 4$. Then, $\mathcal{D} \subseteq \mathcal{D}^{PC}$.*

$$\phi = \begin{array}{c} \\ \\ \\ \end{array}\begin{array}{cc} \begin{array}{ccc} a & b & c \end{array} & \\ \begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{pmatrix} & \begin{array}{c} a \\ b \\ c \end{array} \end{array}$$

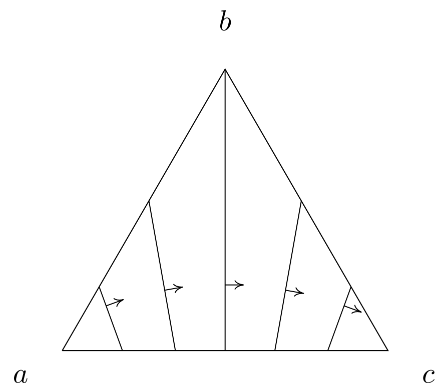

Figure 2: Illustration of preferences based on pairwise comparisons when preferences on pure outcomes are given by the transitive relation $a \succ b \succ c$. The left-hand side shows the corresponding SSB function and the right-hand side the Marschak-Machina probability triangle. The arrows represent the normal vectors to the indifference curves (pointing towards the lower contour set). Each indifference curve separates the corresponding upper and lower contour set.

Theorem 1 is illustrated in Figure 1. Preferences based on pairwise comparisons are quite natural and can be seen as the canonical SSB representation consistent with a given ordinal preference relation over alternatives. For a preference relation $\succ$ that can be represented using an SSB function in Φ^{PC} and two outcomes $p, q \in \Delta$ we have that

$$p \succ q \quad \text{iff} \quad p^T \phi\, q > 0 \quad \text{iff} \quad \sum_{a,b\colon a \succ b} p(a) \cdot q(b) > \sum_{a,b\colon a \succ b} q(a) \cdot p(b).$$

If p and q are interpreted as lotteries, p is preferred to q if and only if p is more likely to return a more preferred alternative than q. Alternatively, the terms in the inequality above can be associated with the probability of *ex ante* regret. Then, p is preferred to q if its choice results in less *ex ante* regret.

Preferences based on pairwise comparisons have previously been considered in decision theory (Blyth, 1972; Packard, 1982; Blavatskyy, 2006). Packard (1982) calls them the *rule of expected dominance* and Blavatskyy (2006) refers to them as a *preference for the most probable winner*. Aziz et al. (2015, 2017) and Brandl et al. (2017) have studied Pareto efficiency, strategyproofness, and related properties with respect to these preferences.

Figure 2 illustrates preferences based on pairwise comparisons for three transitively ordered alternatives.[10] Blavatskyy (2006) gives an axiomatic characterization using Fishburn's SSB axioms and an additional axiom called *fanning-in*, which essentially prescribes that indifference curves are not parallel, but fanning in at a certain rate (see Figure 2).

[10]For three alternatives, preferences based on pairwise comparisons as depicted in Figure 2 can be represented by a WL function with utility function $u(a) = u(b) = 1$ and $u(c) = 0$ and weight function $w(a) = 0$ and $w(b) = w(c) = 1$.

$$\phi = \begin{array}{c} \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 1 \\ -1 & 0 & 1 & 1 \\ -1 & -1 & 0 & 1 \\ -1 & -1 & -1 & 0 \end{pmatrix} \end{array} \begin{array}{c} \\ a \\ b \\ c \\ d \end{array}$$

	a	b	c	d
p	0	0	1	0
q	$2/5$	0	0	$3/5$
r	0	$3/5$	0	$2/5$

Figure 3: Illustration of preferences based on pairwise comparisons when preferences on pure outcomes are given by the transitive relation $a \succ b \succ c \succ d$. The left-hand side shows the corresponding SSB function. The preferences between the three outcomes p, q, and r, defined in the table on the right-hand side, are cyclic: $\phi(p, q) = p^T \phi\, q = 3/5 - 2/5 = 1/5 > 0$, $\phi(q, r) = q^T \phi\, r = 2/5 - (3/5)^2 = 4/25 > 0$, and $\phi(r, p) = r^T \phi\, p = 3/5 - 2/5 = 1/5 > 0$. Hence, $p \succ q \succ r \succ p$.

As a corollary of Theorem 1, fanning-in is implied by Fishburn's SSB axioms and Arrow's axioms. Blavatskyy cites extensive experimental evidence for the fanning-in of indifference curves.

When there are at least four alternatives, preferences based on pairwise comparisons can be cyclic even when preferences over pure outcomes are transitive. This phenomenon, known as the *Steinhaus-Trybula paradox*, is illustrated in Figure 3 (see, e.g., Steinhaus and Trybula, 1959; Blyth, 1972; Packard, 1982; Rubinstein and Segal, 2012; Butler et al., 2016). Butler et al. (2016) have conducted an extensive experimental study of the Steinhaus-Trybula paradox and found significant evidence for preferences based on pairwise comparisons.

6. Characterization of the Social Welfare Function

Theorem 1 has established that anonymous Arrovian aggregation is only possible if individual preferences are based on pairwise comparisons. It turns out that SWFs on this domain satisfy IIA and Pareto indifference if and only if outcomes are compared based on a linear combination of the individual SSB functions. Coefficients of the individual SSB functions may be zero or even negative. This, for example, allows for dictatorial SWFs where the collective preference is identical to the preference relation of one pre-determined agent. When assuming full Pareto optimality, the coefficients assigned to these SSB functions have to be positive, which rules out dictatorial SWFs.

Theorem 2. *Let f be an Arrovian SWF on some domain $\mathcal{D} \subseteq \mathcal{D}^{PC}$ with $|U| \geq 5$. Then, there are $w_1, \dots, w_n \in \mathbb{R}_{>0}$ such that*

$$f(R) \equiv \sum_{i \in N} w_i \phi_i \text{ for all } R \in \mathcal{D}^N.$$

$$\phi = \begin{array}{c} \\ \\ \\ \end{array}\begin{array}{ccc} a & b & c \\ \end{array}$$

$$\begin{array}{cc} & \begin{array}{ccc} a & b & c \end{array} \\ \phi = & \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \begin{array}{c} a \\ b \\ c \end{array} \end{array}$$

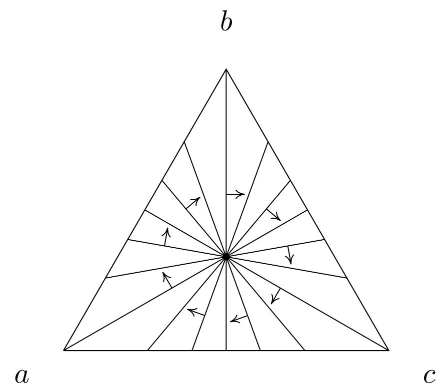

Figure 4: Illustration of collective preferences returned by the unique anonymous Arrovian SWF in the case of Condorcet's paradox. The left-hand side shows the collective SSB function and the right-hand side the Marschak-Machina probability triangle. The arrows represent the normal vectors to the indifference curves (pointing towards the lower contour set). Each indifference curve separates the corresponding upper and lower contour set. The unique most preferred outcome is $1/3\, a + 1/3\, b + 1/3\, c$.

This can be seen as a multi-profile version of Harsanyi's Social Aggregation Theorem (see Section 2) for SSB utilities, where IIA allows us to connect coefficients across different profiles. When furthermore assuming anonymity, the coefficients of all SSB functions have to be identical and we obtain the following complete characterization.

Corollary 1. *Let f be an anonymous Arrovian SWF. Then,*

$$f(R) \equiv \sum_{i \in N} \phi_i \text{ for all } R \in \mathcal{D}^N.$$

The unique anonymous Arrovian SWF is computationally tractable: two outcomes can be compared by straightforward matrix-vector multiplications while a maximal outcome can be found using linear programming. For illustrative purposes, consider the classic Condorcet example where there are three agents with the following transitive preferences over pure outcomes: $a \succ_1 b \succ_1 c$, $b \succ_2 c \succ_2 a$, and $c \succ_3 a \succ_3 b$. These preferences are represented by $\phi_1, \phi_2, \phi_3 \in \Phi^{PC}$ where

$$\begin{array}{cc} & \begin{array}{ccc} a & b & c \end{array} \\ \phi_1 = & \begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{pmatrix} \begin{array}{c} a \\ b \\ c \end{array} \end{array}, \quad \begin{array}{cc} & \begin{array}{ccc} a & b & c \end{array} \\ \phi_2 = & \begin{pmatrix} 0 & -1 & -1 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \begin{array}{c} a \\ b \\ c \end{array} \end{array}, \text{ and } \quad \begin{array}{cc} & \begin{array}{ccc} a & b & c \end{array} \\ \phi_3 = & \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & -1 \\ 1 & 1 & 0 \end{pmatrix} \begin{array}{c} a \\ b \\ c \end{array} \end{array}.$$

Note that the pairwise majority relation is cyclic, since there are majorities for a over b, b over c, and c over a. The unique anonymous Arrovian SWF f aggregates preferences by adding the individual utility representations, i.e.,

$$f(R) \equiv \sum_{i\in N} \phi_i = \begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \end{array} \begin{array}{c} \\ a \\ b \\ c \end{array} .$$

Figure 4 shows the collective preference relation represented by this matrix. The unique most preferred outcome is $1/3\, a + 1/3\, b + 1/3\, c$.[11]

7. Discussion

Our results challenge the traditional—transitive—way of thinking about individual and collective preferences, which has been largely influenced by the pervasiveness of scores and grades. Once one accepts that individual preferences may be intransitive, Pareto optimality implies the same for collective preferences. While such a collective welfare relation does not provide a ranking of all possible outcomes, it nevertheless allows for the comparison of arbitrary pairs of outcomes and identifies maximal (and minimal) elements in each feasible set of outcomes.[12] A compelling opinion on transitivity, which matches the narrative of this paper, is expressed in the following quote by decision theorist Peter Fishburn:

> Transitivity is obviously a great practical convenience and a nice thing to have for mathematical purposes, but long ago this author ceased to understand why it should be a cornerstone of normative decision theory. [...] The presence of intransitive preferences complicates matters [...] however, it is not cause enough to reject intransitivity. An analogous rejection of non-Euclidean geometry in physics would have kept the familiar and simpler Newtonian mechanics in place, but that was not to be. Indeed, intransitivity challenges us to consider more flexible models that retain as much simplicity and elegance as circumstances allow. It challenges old ways of analyzing decisions and suggests new possibilities. (Fishburn, 1991, pp. 115–117)

Theorem 2, the main result of this paper, can be viewed as an intermediary between Harsanyi's Social Aggregation Theorem and Arrow's Impossibility Theorem: it uses Arrow's axioms to derive Harsanyi's utilitarian consequence. Clearly, the form of utilitarianism characterized in Theorem 2 is rather restricted as, due to Theorem 1, it does not

[11] This outcome represents a somewhat unusual unique maximal element because it is not *strictly* preferred to any of the other outcomes. This is due to the contrived nature of the example and only happens if the support of a maximal outcome contains all alternatives.

[12] Bernheim and Rangel (2009) have recently also put forward a relaxed—intransitive—notion of welfare and defended it as "a viable welfare criterion" because it guarantees the existence of maximal elements for finite sets. In fact, Bernheim and Rangel write that "to conduct useful welfare analysis, one does not require transitivity" (see also Bernheim, 2009).

allow for intensities of individual preferences.[13] In fact, it is no more "utilitarian" than approval voting or Borda's rule, which are also based on the summation of scores in purely ordinal contexts. In contrast to these rules, however, pairwise utilitarianism respects majority rule and thereby reconciles Borda's and Condorcet's seemingly conflicting views on preference aggregation (see, e.g., Black, 1958; Young, 1988, 1995). Theorem 2 entails that Arrow's axioms rule out intensities of individual preferences, but do allow for—and in fact require—intensities of collective preferences.

Finally, we would like to close this section with a remarkable quote by Kenneth Arrow, in which he draws the readers attention precisely to the avenue pursued in this paper.

> It seems that the essential point is, and this is of general bearing, that, if conceptually we imagine a choice being made between two alternatives, we cannot exclude any probability distribution over those two choices as a possible alternative. The precise shape of a formulation of rationality which takes the last point into account or the consequences of such a reformulation on the theory of choice in general or the theory of social choice in particular cannot be foreseen; but it is at least a possibility, to which attention should be drawn, that the paradox to be discussed below might be resolved by such a broader concept of rationality [...] Many writers have felt that the assumption of rationality, in the sense of a one-dimensional ordering of all possible alternatives, is absolutely necessary for economic theorizing [...] There seems to be no logical necessity for this viewpoint; we could just as well build up our economic theory on other assumptions as to the structure of choice functions if the facts seemed to call for it. (Arrow, 1951, pp. 20–21)

8. Remarks

We conclude the paper with a number of technical remarks.

Remark 1 (Transitivity). When also requiring transitivity of individual preferences, our result immediately turns into an impossibility, which follows from Theorem 1 and the example given in Figure 3. This implies the impossibility of anonymous Arrovian aggregation of WL preferences (and thereby of vNM preferences), even when collective preferences need not be transitive.[14]

[13] One may even question whether this form of preference aggregation really qualifies as *utilitarianism*. However, in a similar vein, one could also question whether Harsanyi's Social Aggregation Theorem leads to a form of utilitarianism that allows for preference intensities because vNM utilities are merely a compact representation of ordinal preferences over lotteries.

[14] When collective preferences have to be transitive as well, this impossibility directly follows from Arrow's theorem by only considering pure outcomes. IIA and Pareto optimality only become weaker while non-dictatorship is strengthened (dictators only need to be able to dictate strict preferences over pure outcomes). The latter is implied by anonymity.

Remark 2 (Anonymity). Theorem 1 does not hold without assuming anonymity. Let $U = \{a, b, c, d\}$ and $N = \{1, 2, 3\}$ and consider the SSB function

$$\phi = \begin{pmatrix} 0 & 1 & 1 & 1+\varepsilon \\ -1 & 0 & 1 & 1 \\ -1 & -1 & 0 & 1 \\ -(1+\varepsilon) & -1 & -1 & 0 \end{pmatrix}$$

for some $\varepsilon \in (0, 1/4)$. Let $\mathcal{D} = \mathcal{D}^{PC} \cup \{\succ\, \in \mathcal{R} \colon \succ\, \equiv \phi^\pi$ for some $\pi \in \Pi(U)\}$. Then $\mathcal{D}$ satisfies all our domain assumptions (cf. Section 4). The SWF $f \colon \mathcal{D} \to \mathcal{R}$, $f(R) \equiv 2\phi_1 + 3\phi_2 + 4\phi_3$ satisfies IIA and Pareto optimality but violates anonymity. Note that f is not dictatorial. Hence, Theorem 1 does not hold when weakening anonymity to non-dictatorship.

Remark 3 (Symmetry). Theorem 1 also holds when collective preferences are not required to satisfy the symmetry axiom. Whether symmetry is required for individual preferences in Theorem 1 and for collective preferences in Theorem 2 is open.

Remark 4 (Tightness of Bounds). Theorem 1 does not hold if $|U| < 4$, which is the same bound as for the result by Kalai and Schmeidler (1977a). This stems from the fact that for $U = \{a, b, c\}$, IIA only has non-trivial implications for feasible sets of the form $\Delta_{\{a,b\}}$ for some $a, b \in U$. For every possible preference over a and b, there is exactly one continuous and convex preference relation on $\Delta_{\{a,b\}}$ consistent with it. Hence, IIA only has non-trivial implications for the collective preferences over pure outcomes. However, even for three alternatives, the domains of preferences which allow for anonymous Arrovian aggregation are severely restricted. In particular, Lemmas 1 to 4 still hold. Any such domain contains exactly one SSB function ϕ for every strict order over U, which takes the form

$$\phi = \begin{pmatrix} 0 & 1 & \lambda \\ -1 & 0 & 1 \\ -\lambda & -1 & 0 \end{pmatrix}$$

for some $\lambda \in \mathbb{R}_{>0}$ that is fixed across all strict orders. For $1 < \lambda < 1 + 1/n$, affine utilitarianism with weight 1 for all agents constitutes an Arrovian SWF on the corresponding domain.

Theorem 2 does not hold if $|U| < 5$. Let $U = \{a, b, c, d\}$, $\mathcal{D} = \mathcal{D}^{PC}$, and $\hat{R} \equiv (\phi_1, \phi_2, \phi_3, \phi_4, \dots)$ such that every SSB function in $\mathcal{D} \setminus \{0\}$ appears exactly once in the preferences of the agents in $N \setminus \{1, 2, 3, 4\}$ and

$$\phi_1 = \phi_2 = \begin{pmatrix} 0 & 1 & 1 & 1 \\ -1 & 0 & 1 & 1 \\ -1 & -1 & 0 & 1 \\ -1 & -1 & -1 & 0 \end{pmatrix}, \phi_3 = \begin{pmatrix} 0 & 1 & -1 & -1 \\ -1 & 0 & -1 & -1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & -1 & 0 \end{pmatrix}, \phi_4 = \begin{pmatrix} 0 & 1 & -1 & -1 \\ -1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

Then, Pareto optimality has no implications for $\hat{R}$. Let $f\colon \mathcal{D}^N \to \Phi$, $f(R) \equiv \sum_{i\in N} \phi_i$ except that

$$f(\hat{R}) \equiv \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}.$$

f satisfies Pareto optimality and IIA. The proof of Theorem 2 fails at Lemma 7.

Remark 5 (Domain Conditions). In Section 4, we specified domain richness conditions used in our proofs. The last condition required that every transitive and asymmetric relation on pure outcomes is represented by at least one relation in $\mathcal{D}$. For Theorems 1 and 2, this condition is only required for every transitive and asymmetric relation on *four* and *five* pure outcomes, respectively. Furthermore, to derive the conclusion of Theorem 1, a weaker condition suffices: if $\succ \in \mathcal{D}$ with $a \succ b \succ c$ and $a \succ c$ for some $a, b, c \in U$, then there is some $\succ' \in \mathcal{D}$ with $a \succ' b \succ' c \succ' x$ and $a \succ' c$ for some $x \in U$. This condition also covers the domain of dichotomous preferences.

Remark 6 (Infinite Universes). Fishburn (1984c) shows that under additional technical assumptions about the measure space and $\succ$, the SSB representation holds for probability measures over arbitrary (possibly infinite) sets of alternatives. Our results extend to this framework without modifications to the proofs.

Acknowledgements

This material is based on work supported by the Deutsche Forschungsgemeinschaft under grant BR 2312/10-1. Early versions of this paper were presented at the Yale Microeconomic Theory Reading Group (September, 2016) and the 5th Annual D-TEA Workshop in Paris (May, 2017). The authors thank Tilman Börgers, Paul Harrenstein, Ehud Kalai, Michel Le Breton, Philippe Mongin, Klaus Nehring, Dominik Peters, Mirek Pištěk, Marcus Pivato, Clemens Puppe, Larry Samuelson, and Gerelt Tserenjigmid for helpful comments.

References

M. Allais. Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine. *Econometrica*, 21(4):503–546, 1953.

P. Anand. The philosophy of intransitive preference. *The Economic Journal*, 103(417):337–346, 1993.

P. Anand. Rationality and intransitive preference: Foundations for the modern view. In P. Anand, P. K. Pattanaik, and C. Puppe, editors, *The Handbook of Rational and Social Choice*, chapter 6. Oxford University Press, 2009.

K. J. Arrow. *Social Choice and Individual Values*. New Haven: Cowles Foundation, 1st edition, 1951. 2nd edition 1963.

K. J. Arrow. Rational choice functions and orderings. *Economica*, 26:121–127, 1959.

H. Aziz, F. Brandt, and P. Stursberg. On popular random assignments. In *Proceedings of the 6th International Symposium on Algorithmic Game Theory (SAGT)*, volume 8146 of *Lecture Notes in Computer Science (LNCS)*, pages 183–194. Springer-Verlag, 2013.

H. Aziz, F. Brandl, and F. Brandt. Universal Pareto dominance and welfare for plausible utility functions. *Journal of Mathematical Economics*, 60:123–133, 2015.

H. Aziz, F. Brandl, F. Brandt, and M. Brill. On the tradeoff between efficiency and strategyproofness. 2017. Working paper.

J. S. Banks. Acyclic social choice from finite sets. *Social Choice and Welfare*, 12:293–310, 1995.

M. Bar-Hillel and A. Margalit. How vicious are cycles of intransitive choice? *Theory and Decision*, 24(2):119–145, 1988.

T. C. Bergstrom. When non-transitive relations take maxima and competitive equilibrium can't be beat. In W. Neuefeind and R. G. Riezmann, editors, *Economic Theory and International Trade (Essays in Memoriam of J. Trout Rader)*, pages 29–52. Springer-Verlag, 1992.

B. D. Bernheim. Behavioral welfare economics. *Journal of the European Economic Association*, 7 (2–3):267–319, 2009.

B. D. Bernheim and A. Rangel. Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics*, 124(1):51–104, 2009.

D. Black. On the rationale of group decision-making. *Journal of Political Economy*, 56(1):23–34, 1948.

D. Black. *The Theory of Committees and Elections*. Cambridge University Press, 1958.

D. H. Blair and R. A. Pollack. Polychromatic acyclic tours in colored multigraphs. *Mathematics of Operations Research*, 8(3):471–476, 1983.

D. H. Blair and R. A. Pollak. Acyclic collective choice rules. *Econometrica*, 50(4):931–943, 1982.

J. H. Blau and R. Deb. Social decision functions and the veto. *Econometrica*, 45(4):871–879, 1977.

P. R. Blavatskyy. Axiomatization of a preference for most probable winner. *Theory and Decision*, 60(1):17–33, 2006.

C. R. Blyth. Some probability paradoxes in choice from among random alternatives. *Journal of the American Statistical Association*, 67(338):366–373, 1972.

K. C. Border. Social welfare functions for economic environments with and without the Pareto principle. *Journal of Economic Theory*, 29(2):205–216, 1983.

G. Bordes and M. Le Breton. Arrovian theorems with private alternatives domains and selfish individuals. *Journal of Economic Theory*, 47(2):257–281, 1989.

G. Bordes and M. Le Breton. Arrovian theorems for economic domains: Assignments, matchings and pairings. *Social Choice and Welfare*, 7(3):193–208, 1990a.

G. Bordes and M. Le Breton. Arrovian theorems for economic domains. the case where there are simultaneously private and public goods. *Social Choice and Welfare*, 7(1):1–17, 1990b.

T. Börgers and Y.-M. Choo. Revealed relative utilitarianism. Working paper, 2015.

T. Börgers and Y.-M. Choo. A counterexample to Dhillon (1998). *Social Choice and Welfare*, 48 (4):837–843, 2017.

F. Brandl, F. Brandt, and H. G. Seedig. Consistent probabilistic social choice. *Econometrica*, 84 (5):1839–1880, 2016.

F. Brandl, F. Brandt, and J. Hofbauer. Welfare maximization entices participation. 2017. Working paper.

F. Brandt and P. Harrenstein. Set-rationalizable choice and self-stability. *Journal of Economic Theory*, 146(4):1721–1731, 2011.

F. Brandt, M. Brill, H. G. Seedig, and W. Suksompong. On the structure of stable tournament solutions. *Economic Theory*, 2017. Forthcoming.

D. J. Brown. Aggregation of preferences. *Quarterly Journal of Economics*, 89(3):456–469, 1975.

D. Butler, G. Pogrebna, and P. R. Blavatskyy. Predictably intransitive preferences. 2016. Working paper.

D. E. Campbell. Arrow's theorem for economic environments and effective social preferences. *Social Choice and Welfare*, 6(4):325–329, 1989.

D. E. Campbell and J. S. Kelly. Impossibility theorems in the Arrovian framework. In K. J. Arrow, A. K. Sen, and K. Suzumura, editors, *Handbook of Social Choice and Welfare*, volume 1, chapter 1. Elsevier, 2002.

S. H. Chew. A generalization of the quasilinear mean with applications to the measurement of income and decision theory resolving the Allais paradox. *Econometrica*, 51(4):1065–1092, 1983.

M. Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Imprimerie Royale, 1785. Facsimile published in 1972 by Chelsea Publishing Company, New York.

P. Dasgupta and E. Maskin. On the robustness of majority rule. *Journal of the European Economic Association*, 6(5):949–973, 2008.

C. d'Aspremont and L. Gevers. Social welfare functionals and interpersonal comparability. In K. J. Arrow, A. K. Sen, and K. Suzumura, editors, *Handbook of Social Choice and Welfare*, volume 1, chapter 10, pages 459–541. Elsevier, 2002.

A. Dhillon. Extended Pareto rules and relative utilitarianism. *Social Choice and Welfare*, 15(4): 521–542, 1998.

A. Dhillon and J.-F. Mertens. Relative utilitarianism. *Econometrica*, 67(3):471–498, 1999.

L. Ehlers and T. Storcken. Arrow's possibility theorem for one-dimensional single-peaked preferences. *Games and Economic Behavior*, 64(2):533–547, 2008.

P. C. Fishburn. The irrationality of transitity in social choice. *Behavioral Science*, 15:119–123, 1970.

P. C. Fishburn. Nontransitive measurable utility. *Journal of Mathematical Psychology*, 26(1):31–67, 1982.

P. C. Fishburn. Transitive measurable utility. *Journal of Economic Theory*, 31(2):293–317, 1983.

P. C. Fishburn. Probabilistic social choice based on simple voting comparisons. *Review of Economic Studies*, 51(4):683–692, 1984a.

P. C. Fishburn. SSB utility theory: An economic perspective. *Mathematical Social Sciences*, 8(1): 63–94, 1984b.

P. C. Fishburn. Dominance in SSB utility theory. *Journal of Economic Theory*, 34(1):130–148, 1984c.

P. C. Fishburn. *Nonlinear preference and utility theory*. The Johns Hopkins University Press, 1988.

P. C. Fishburn. Nontransitive preferences in decision theory. *Journal of Risk and Uncertainty*, 4 (2):113–134, 1991.

P. C. Fishburn and W. V. Gehrlein. Aggregation theory for SSB utility functionals. *Journal of Economic Theory*, 42(2):352–369, 1987.

P. C. Fishburn and P. Wakker. The invention of the independence condition for preferences. *Management Science*, 41(7):1130–1144, 1995.

M. Fleurbaey, M. Salles, and J. A. Weymark, editors. *Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls*. Cambridge University Press, 2008.

D. M. Grether and C. R. Plott. Economic theory of choice and the preference reversal phenomenon. *American Economic Review*, 69(4):623–638, 1979.

J. C. Harsanyi. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63(4):309–321, 1955.

A. Hylland. Aggregation procedure for cardinal preferences: A comment. *Econometrica*, 48(2): 539–542, 1980.

K. Inada. The simple majority decision rule. *Econometrica*, 37:490–506, 1969.

D. Kahnemann and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292, 1979.

E. Kalai and D. Schmeidler. An admissible set occurring in various bargaining situations. *Journal of Economic Theory*, 14:402–411, 1977a.

E. Kalai and D. Schmeidler. Aggregation procedure for cardinal preferences: A formulation and proof of Samuelson's impossibility conjecture. *Econometrica*, 45(6):1431–1438, 1977b.

E. Kalai, E. Muller, and M. A. Satterthwaite. Social welfare functions when preferences are convex, strictly monotonic, and continuous. *Public Choice*, 34(1):87–97, 1979.

T. Kavitha, J. Mestre, and M. Nasre. Popular mixed matchings. *Theoretical Computer Science*, 412(24):2679–2690, 2011.

J. S. Kelly. *Arrow Impossibility Theorems*. Economic Theory and Mathematical Economics. Academic Press, 1978.

G. Kreweras. Aggregation of preference orderings. In *Mathematics and Social Sciences I: Proceedings of the seminars of Menthon-Saint-Bernard, France (1–27 July 1960) and of Gösing, Austria (3–27 July 1962)*, pages 73–79, 1965.

G. Laffond, J.-F. Laslier, and M. Le Breton. The bipartisan set of a tournament game. *Games and Economic Behavior*, 5(1):182–201, 1993.

J.-F. Laslier. *Tournament Solutions and Majority Voting*. Springer-Verlag, 1997.

J.-F. Laslier. Aggregation of preferences with a variable set of alternatives. *Social Choice and Welfare*, 17(2):269–282, 2000.

M. Le Breton. *Essais sur les fondements de l'analyse économique de l'inégalité*. Thèse pour le doctorat d'État en sciences Économiques, Université de Rennes 1, 1986.

M. Le Breton and J. A. Weymark. Arrovian social choice theory on economic domains. In K. J. Arrow, A. K. Sen, and K. Suzumura, editors, *Handbook of Social Choice and Welfare*, volume 2, chapter 17. North-Holland, 2011.

J.-V. Llinares. Unified treatment of the problem of existence of maximal elements in binary relations: A characterization. *Journal of Mathematical Economics*, 29(3):285–302, 1998.

G. Loomes and R. Sugden. Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 92(368):805–824, 1982.

G. Loomes and R. Sugden. Some implications of a more general form of regret theory. *Journal of Economic Theory*, 41(2):270–287, 1987.

M. J. Machina. Generalized expected utility analysis and the nature of observed violations of the independence axiom. In B. Stigum and F. Wenstop, editors, *Foundations of Utility and Risk Theory with Applications*, chapter 5. Springer, 1983.

M. J. Machina. Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature*, 27(4):1622–1668, 1989.

A. Mas-Colell and H. Sonnenschein. General possibility theorems for group decisions. *Review of Economic Studies*, 39(2):185–192, 1972.

E. Maskin and A. Sen. *The Arrow Impossiblity Theorem.* Columbia University Press, 2014.

K. May. Intransitivity, utility, and the aggregation of preference patters. *Econometrica*, 22(1):1–13, 1954.

E. F. McClennen. Sure-thing doubts. In P. Gärdenfors and N.-E. Sahlin, editors, *Decision, Probability and Utility*, chapter 10. Cambridge University Press, 1988.

P. Mongin. Harsanyi's aggregation theorem: multi-profile version and unsettled questions. *Social Choice and Welfare*, 11(4):331–354, 1994.

H. Nishimura. The transitive core: Inference of welfare from nontransitive preference relations. *Theoretical Economics*, 2017. Forthcoming.

D. J. Packard. Cyclical preference logic. *Theory and Decision*, 14(4):415–426, 1982.

J. Redekop. Arrow theorems in econonmic environments. In W. A. Barnett, H. Moulin, M. Salles, and N. J. Schofield, editors, *Social choice, welfare, and ethics*, pages 163–185. Cambridge University Press, 1995.

A. Rubinstein and U. Segal. On the likelihood of cyclic comparisons. *Journal of Economic Theory*, 147(6):2483–2491, 2012.

P. A. Samuelson. A note on the pure theory of consumers' behaviour. *Econometrica*, 5:61–71, 1938.

P. A. Samuelson. Arrow's mathematical politics. In S. Hook, editor, *Human Values and Economic Policy*, pages 41–51. New York University Press, 1967.

T. Schwartz. *The Logic of Collective Choice.* Columbia University Press, 1986.

A. K. Sen. Quasi-transitivity, rational choice and collective decision. *Review of Economic Studies*, 36(3):381–393, 1969.

A. K. Sen. *Collective Choice and Social Welfare.* North-Holland, 1970.

A. K. Sen. Choice functions and revealed preference. *Review of Economic Studies*, 38(3):307–317, 1971.

A. K. Sen. Social choice theory: A re-examination. *Econometrica*, 45(1):53–89, 1977.

A. K. Sen. Social choice theory. In K. J. Arrow and M. D. Intriligator, editors, *Handbook of Mathematical Economics*, volume 3, chapter 22, pages 1073–1181. Elsevier, 1986.

A. K. Sen. Rationality and social choice. *American Economic Review*, 85(1):1–24, 1995.

A. K. Sen. *Collective Choice and Social Welfare.* Penguin, expanded edition, 2017.

A. K. Sen and P. K. Pattanaik. Necessary and sufficient conditions for rational choice under majority decision. *Journal of Economic Theory*, 1:178–202, 1969.

H. Sonnenschein. Demand theory without transitive preference with applications to the theory of competitive equilibrium. In J. Chipman, L. Hurwicz, M. Richter, and H. Sonnenschein, editors, *Preferences, Utility and Demand.* Houghton Mifflin Harcourt, 1971.

H. Steinhaus and S. Trybula. On a paradox in applied probabilities. *Bulletin of the Polish Academy of Sciences*, 7:67–69, 1959.

A. H. Turunen-Red and J. A. Weymark. Linear aggregation of SSB utility functionals. *Theory and Decision*, 46(3):281–294, 1999.

A. Tversky. Intransivitiy of preferences. *Psychological Review*, 76:31–48, 1969.

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 2nd edition, 1947.

H. P. Young. Condorcet's theory of voting. *The American Political Science Review*, 82(4):1231–1244, 1988.

H. P. Young. Optimal voting rules. *Journal of Economic Perspectives*, 9(1):51–64, 1995.

APPENDIX

A. Arrovian Impossibilities for vNM Preferences

As mentioned in Section 2, there are a number of Arrovian impossibilities when preferences over lotteries satisfy the von Neumann-Morgenstern (vNM) axioms and thus can be represented by assigning cardinal utilities to alternatives such that lotteries are compared based on the expected utility they produce. We believe that a detailed comparison of these results which have appeared in different branches of social choice theory and welfare economics is in order.

The literature on *economic domains* uses a framework very similar to the one studied in this paper (see Le Breton and Weymark, 2011). A key question is whether Arrow's impossibility remains intact if the domain of admissible preference profiles is subject to certain structural restrictions. Many results in this area rely on the local approach due to Kalai et al. (1979), who proposed a simple domain condition that is sufficient for Arrow's impossibility. Le Breton (1986) has shown that this condition is satisfied by the domain of vNM preferences, which implies Arrow's impossibility (see also Le Breton and Weymark (2011, p. 214)). The corresponding IIA condition is defined for arbitrary pairs of lotteries, or—equivalently—arbitrary feasible sets of size two (which implies IIA for arbitrary feasible sets of lotteries). In view of the structure of the set of lotteries, weaker IIA conditions (for example, restricted to convex feasible sets) seem natural.

Sen (1970) has initiated the study of so-called *social welfare functionals (SWFLs)*, which map a profile of cardinal utilities to a transitive and complete collective preference relation (see also d'Aspremont and Gevers, 2002). The definitions of IIA and Pareto optimality can be straightforwardly extended to SWFLs. Note, however, that IIA takes into account the absolute values of utilities (rather than only ordinal comparisons between these values). This allows for Pareto optimal SWFLs that satisfy IIA, for example by adding individual utilities (utilitarianism).

vNM utilities are invariant under positive affine transformations. To account for this, Sen (1970) introduced the axiom of *cardinality and non-comparability*, which prescribes that collective preferences returned by the SWFL are invariant under positive affine transformations of the individual utility functions. However, this assumption effectively turns the problem into a problem of ordinal preference aggregation because the utility values assigned to two different alternatives in two different utility profiles can be made identical across profiles by applying a positive affine transformation. Hence, IIA implies an ordinal version of IIA which only takes into account the ordinal comparisons between utility values and Arrow's original theorem holds (Sen, 1970, Theorem 8*2).

There are two ways to interpret this result. First, one can view the set of alternatives as the set consisting of only degenerate lotteries. This leads to weak notions of Pareto optimality and IIA because they are only concerned with degenerate lotteries. Non-dictatorship, on the other hand, becomes much stronger because a dictator can only enforce his (strict) preferences over degenerate lotteries, rather than all lotteries. Alternatively, one can define the set of alternatives as the set of all lotteries. This gives rise to stronger notions of Pareto optimality and IIA based on pairs of lotteries, rather than pairs of degenerate lotteries. In this model, non-dictatorship is defined by excluding agents who can enforce their (strict) preferences over lotteries. Like Arrow's theorem, Sen's result assumes an unrestricted domain of preferences (or ordinal utilities, respectively). Expected utility functions over a set of lotteries, however, are subject to certain structural constraints (described by the vNM axioms independence and continuity). This gap is filled by Mongin (1994, Proposition 3), who has shown that Sen's result still holds when the set of alternatives is a convex subset of some vector space with mixture-preserving (i.e., affine) utility functions, which includes the domain of lotteries over some finite set of alternatives as a special case.[15] The Pareto condition used by Mongin is identical to the one used in his paper and therefore slightly stronger than the one used by Arrow, Sen, and Le Breton.

A very strong impossibility for vNM preferences was given by Kalai and Schmeidler (1977b) (and later improved by Hylland (1980)). Kalai and Schmeidler consider "cardinal" preference relations represented by equivalence classes of utility functions that can be transformed into each other using positive affine transformations and *cardinal social welfare functions*, which map a profile of cardinal preference relations to a collective cardinal preference relation. The set of alternatives is defined as the set of degenerate lotteries like in the first interpretation of Sen's result above. Preferences over lotteries are implicit in each equivalence class of utility functions. When interpreted in our ordinal framework, they prove an Arrovian impossibility when individual and collective preferences over lotteries are subject to the vNM axioms and there are at least four alternatives. In contrast to the results by Le Breton, Sen, and Mongin, IIA is only required for feasible sets given by the convex combination of degenerate lotteries and non-dictatorship only rules out projections. The theorem thus uses weaker notions of Pareto optimality, IIA, and non-dictatorship at the expense of also requiring the vNM axioms for the collective preference relation. When replacing non-dictatorship with anonymity, our Theorem 1 implies a similar impossibility, even without requiring collective preferences to be transitive (see Remark 1). We use Kalai and Schmeidler's weak IIA notion, but Mongin's strong notion of Pareto optimality.

B. Characterization of the Domain

We first prove a crucial lemma, which shows that continuous and convex preference relations are completely determined by their symmetric part up to orientation. This generalizes Theorem 2 by Fishburn and Gehrlein (1987), who showed the same statement for SSB preferences (i.e., they additionally assume symmetry). A self-contained proof of Lemma 1 is given in Appendix D. The weaker version by Fishburn and Gehrlein is sufficient for our main result, but we believe Lemma 1 may be of independent interest, e.g., when trying to strengthen Theorems 1 and 2.

[15] The vector space is required to be at least of dimension 2, which corresponds to the set of lotteries over at least three degenerate lotteries.

Lemma 1. *Let $\succ, \hat{\succ}$ be continuous and convex preference relations. If $\sim \subseteq \hat{\sim}$, then $\hat{\succ} \in \{\succ, \succ^{-1}, \emptyset\}$.*

The next lemma is reminiscent of what is known as the *field expansion lemma* in traditional proofs of Arrow's theorem (see, e.g. Sen, 1986).[16] Let $f\colon \mathcal{D}^N \to \mathcal{R}$ be an SWF, $G, H \subseteq N$, and $a, b \in U$. We say that (G, H) is decisive for a against b, denoted by $a \; D_{G,H} \; b$, if, for all $R \in \mathcal{D}^N$, $a \succ_i b$ for all $i \in G$, $a \sim_i b$ for all $i \in H$, and $b \succ_i a$ for all $i \in N \setminus (G \cup H)$ implies $a \succ b$. Hence, $D_{G,H}$ is a relation on U.

Lemma 2. *Let f be an Arrovian SWF on some domain $\mathcal{D}$, $G, H \subseteq N$, and $a, b \in U$. Then $a \; D_{G,H} \; b$ implies that $D_{G,H} = U \times U$.*

Proof. First we show that $a \; D_{G,H} \; x$ and $b \; D_{G,H} \; x$ for all $x \in U \setminus \{a, b\}$. To this end, let $x \in U \setminus \{a, b\}$ and $\succ_x \in \mathcal{D}$ such that $a \succ_x b \succ_x x$ and $a \succ_x x$ and consider the preference profile

$$R = (\underbrace{\succ_x, \dots, \succ_x}_{G}, \underbrace{\emptyset, \dots, \emptyset}_{H}, \succ_x^{-1}, \dots, \succ_x^{-1}).$$

Since $\succ_x \cap \succ_x^{-1} = \emptyset$, it follows from Pareto indifference and Lemma 1 that $\succ = f(R) \in \{\succ_x, \succ_x^{-1}, \emptyset\}$. Since $a \; D_{G,H} \; b$, $\succ = \succ_x$ remains as the only possibility. Hence, $a \succ x$ and $b \succ x$. By IIA, it follows that $a \; D_{G,H} \; x$ and $b \; D_{G,H} \; x$.

Repeated application of the second statement implies that $D_{G,H}$ is a complete relation. To show that $D_{G,H}$ is symmetric, let $x, y, z \in U$ such that $x \; D_{G,H} \; y$. The first part of the statement implies that $x \; D_{G,H} \; z$. Two applications of the second part of the statement yield $z \; D_{G,H} \; y$ and $y \; D_{G,H} \; x$. Hence, $D_{G,H} = U \times U$. □

Now we show that anonymous Arrovian aggregation is only possible on domains in which preferences over outcomes are completely determined by preferences over pure outcomes.

Lemma 3. *Let f be an anonymous Arrovian SWF on some domain $\mathcal{D}$. Then, $\succ|_A = \hat{\succ}|_A$ implies $\succ|_{\Delta A} = \hat{\succ}|_{\Delta A}$ for all $\succ, \hat{\succ} \in \mathcal{D}$ and $A \in \mathcal{A}$.*

Proof. Let $\succ_0, \hat{\succ}_0 \in \mathcal{D}$ and $A \in \mathcal{A}$ such that $\succ_0|_A = \hat{\succ}_0|_A$. Consider the preference profile

$$R = (\succ_0, \hat{\succ}_0^{-1}, \emptyset, \dots, \emptyset).$$

Note that $R \in \mathcal{D}^N$ since $\mathcal{D}$ satisfies our richness assumptions. Now let $a, b \in U$ and define $\bar{R} = R_{(12)}$ to be identical to R except that the preferences of agents 1 and 2 are exchanged. Anonymity of f implies that $\bar{\succ} = f(\bar{R}) = f(R) = \succ$. Assume for contradiction that $a \succ b$. Then, by IIA, $(\{1\}, N \setminus \{1, 2\})$ is decisive for a against b. Lemma 2 implies that $(\{1\}, N \setminus \{1, 2\})$ is also decisive for b against a. Hence $b \mathrel{\bar{\succ}} a$, which contradicts $\bar{\succ} = \succ$. Thus, $a \sim b$. Since $\succ$ satisfies convexity, we get that $\succ|_{\Delta A} = \emptyset$. If $\succ_0|_{\Delta A} \neq \hat{\succ}_0|_{\Delta A}$, there are $p, q \in \Delta A$ such that $p \succ_0 q$ and not $p \mathrel{\hat{\succ}_0} q$, i.e., $p \mathrel{\hat{\succsim}_0^{-1}} q$. The strict part of Pareto optimality of f implies that $p \succ q$. This contradicts $\succ|_{\Delta A} = \emptyset$. Hence, $\succ_0|_{\Delta A} = \hat{\succ}_0|_{\Delta A}$. □

[16]In contrast to Lemma 2, the consequence of the original field expansion lemma uses a stronger notion of decisiveness.

Lemma 3 is the only part of the proof of Theorem 1 that requires anonymity. A much weaker condition would also suffice: there has to be $R \in \mathcal{D}^N$, $p, q \in \Delta$, $i \in N$, and $f(R) = \succ$ such that $p \succ_i q$ and $p \sim q$.

Next, we show that intensities of preferences between pure outcomes have to be identical.

Lemma 4. *Let f be an anonymous Arrovian SWF on some domain $\mathcal{D}$. Then, for all $\succ_0 \in \mathcal{D}$ and $a, b, c \in U$ with $a \succ_0 b$,*

(i) $b \succ_0 c$ implies $\phi_0(a, b) = \phi_0(b, c)$,

(ii) $a \succ_0 c$ implies $\phi_0(a, b) = \phi_0(a, c)$, and

(iii) $c \succ_0 b$ implies $\phi_0(a, b) = \phi_0(c, b)$.

Proof. Ad (i): continuity implies that $b \sim_0 \lambda a + (1 - \lambda)c$ for some $\lambda \in (0, 1)$. Observe that $\succ_0^{(a,c)}|_{\{a,b,c\}} = \succ_0^{-1}|_{\{a,b,c\}}$. Lemma 3 implies that $\succ_0^{(a,c)}|_{\Delta_{\{a,b,c\}}} = \succ_0^{-1}|_{\Delta_{\{a,b,c\}}}$. Hence, we have $b \sim_0 (1 - \lambda)a + \lambda c$. Convexity of $\succ_0$ then implies that $b \sim_0 1/2\, a + 1/2\, c$. This is equivalent to $\phi_0(a, b) = \phi_0(b, c)$.

Ad (ii): we distinguish two cases.

Case 1 ($b \sim_0 c$): let $i, j \in N$ and consider the preference profile

$$R = (\succ_0, (\succ_0^{(bc)})^{-1}, \emptyset, \dots, \emptyset).$$

As in the proof of Lemma 3, we get that $\succ|_{\Delta_{\{a,b,c\}}} = \emptyset$. Without loss of generality, assume that $\phi_0(a, b) = 1$ and $\phi_0(a, c) = \lambda$ for some $\lambda \in (0, 1]$. Let $p = 1/2\, a + 1/2\, c$ and $q = 1/2\, a + 1/2\, b$. Then $\phi_1(p, q) = \phi_2(p, q) = 1/4\,(1 - \lambda)$. If $\lambda < 1$, the strict part of Pareto optimality of f implies that $p \succ q$. This contradicts $\succ|_{\{a,b,c\}} = \emptyset$. Hence, $\lambda = 1$.

Case 2 ($b \succ_0 c$): assume without loss of generality that $\phi_0(a, b) = 1$. By (i), we get $\phi_0(a, b) = \phi_0(b, c) = 1$. Our richness assumptions on the domain imply that there is $\hat{\succ}_0 \in \mathcal{D}$ with $a \mathrel{\hat{\succ}_0} b \mathrel{\hat{\succ}_0} c$, $a \mathrel{\hat{\succ}_0} c$, and $c \mathrel{\hat{\succ}_0} x$ for some $x \in U$. Lemma 3 implies that $\phi_0|_{\{a,b,c\}} = \hat{\phi}_0|_{\{a,b,c\}}$. Hence, if suffices to show that $\hat{\phi}_0(a, c) = 1$. By (i), we get that $\hat{\phi}_0(a, c) = \hat{\phi}_0(c, x)$ and $\hat{\phi}_0(b, c) = \hat{\phi}_0(c, x) = 1$. Hence, $\hat{\phi}_0(a, c) = 1$.

Ad (iii): the proof is analogous to the proof of (ii). □

Theorem 1. *Let f be an anonymous Arrovian SWF on some domain $\mathcal{D}$. Then $\mathcal{D} \subseteq \mathcal{D}^{PC}$.*

Proof. Let $\succ_0 \in \mathcal{D}$ and $a, b, c, d \in U$ such that $a \succ_0 b$ and $c \succ_0 d$. We have to show that $\phi_0(a, b) = \phi_0(c, d)$. First assume there are $x \in \{a, b\}$ and $y \in \{c, d\}$ such that $x \succ_0 y$ or $y \succ_0 x$. Then, Lemma 4 implies that $\phi_0(a, b) = \phi_0(x, y) = \phi_0(c, d)$ or $\phi_0(a, b) = \phi_0(y, x) = \phi_0(c, d)$, respectively. Otherwise, $x \sim_0 y$ for all $x \in \{a, b\}$ and $y \in \{c, d\}$. This implies that $\succ_0|_{\{a,b,c,d\}} = \succ_0{}^{(ac)(bd)}|_{\{a,b,c,d\}}$. From Lemma 3 it follows that $\succ_0|_{\Delta_{\{a,b,c,d\}}} = \succ_0{}^{(ac)(bd)}|_{\Delta_{\{a,b,c,d\}}}$. Hence, $\phi_0|_{\{a,b,c,d\}} = \phi_0^{(ac)(bd)}|_{\{a,b,c,d\}}$ and $\phi_0(a, b) = \phi_0(c, d)$. □

C. Characterization of the Social Welfare Function

In light of Theorem 1, we will assume throughout this section that $\mathcal{D} \subseteq \mathcal{D}^{PC}$. Except for Theorem 2, all results in this section only require Pareto indifference. Since SSB utilities over outcomes are completely determined by SSB utilities over pure outcomes, we will write ϕ_X instead of the more clumsy ϕ_{Δ_X} for any SSB function ϕ and subset of alternatives $X \subseteq U$.

The following lemmas show that for all preference profiles R and all alternatives a and b, $\phi(a,b)$ only depends on the set of agents who prefer a to b, whenever R is from the domain of PC-preferences and ϕ represents $f(R)$. We first prove that, if an alternative is strictly Pareto dominated, then the intensities of collective preferences between each of the dominating alternatives and the dominated alternative are identical. (Using a symmetric argument, the same can be shown for profiles in which the Pareto dominance is reversed.)

Lemma 5. *Let f be an Arrovian SWF, $a,b,c \in U$, and $R \equiv (\phi_i)_{i\in N}$ such that $\phi_i(a,c) = \phi(b,c) = 1$. Then, $\phi(a,c) = \phi(b,c)$ where $\phi \equiv f(R)$.*

Proof. The idea of the proof is to introduce a fourth alternative, which serves as a calibration device for the intensity of pairwise comparisons, and eventually disregard this alternative using IIA. To this end, let $x \in U$ and consider a preference profile $\hat{R} \in \mathcal{D}^N$ such that $R|_{\{a,b,c\}} = \hat{R}|_{\{a,b,c\}}$ and

$$\hat{R}|_{\{a,b,c,x\}} \equiv (\underbrace{\begin{pmatrix} 0 & & 1 & 1 \\ & 0 & 1 & 1 \\ -1 & -1 & 0 & 1 \\ -1 & -1 & -1 & 0 \end{pmatrix}}_{N}, \dots).$$

The values of $\hat{\phi}_i(a,b)$ for all $i \in N$ are irrelevant.[17] Let $\hat{\phi} \equiv f(\hat{R})$. The Pareto indifference relation with respect to $\hat{R}|_{\{a,c,x\}}$ is identical to $\sim_1|_{\{a,c,x\}}$. The analogous statement holds for the Pareto indifference relation with respect to $\hat{R}|_{\{b,c,x\}}$. Hence, Pareto indifference, Lemma 1, and IIA imply that there are $\alpha, \beta \in \mathbb{R}$ such that

$$\hat{\phi}|_{\{a,c,x\}} = \alpha \begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{pmatrix} \qquad \text{and} \qquad \hat{\phi}|_{\{b,c,x\}} = \beta \begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{pmatrix}.$$

As a consequence, $\alpha = \beta$ and $\hat{\phi}(a,c) = \hat{\phi}(b,c)$. Since $R|_{\{a,b,c\}} = \hat{R}|_{\{a,b,c\}}$, IIA implies that $\phi|_{\{a,b,c\}} \equiv \hat{\phi}|_{\{a,b,c\}}$. Hence, $\phi(a,c) = \phi(b,c)$.[18] □

[17] Also the values $\hat{\phi}_i(x,z)$ for all $z \in \{a,b,c\}$ are irrelevant as long as they are the same for all agents.
[18] Pareto dominance also implies that $\phi(a,c), \phi(b,c) > 0$.

Given a preference profile R, let $N_{ab} = \{i \in N : a \succ_i b\}$ be the set of agents who strictly prefer a over b and $n_{ab} = |N_{ab}|$. Also, let $I_{ab} = N \setminus (N_{ab} \cup N_{ba})$ be the set of agents who are indifferent between a and b.

Lemma 6 shows that for a fixed preference profile, $\phi(a,b)$ only depends on N_{ab} and I_{ab} (and not on the names of the alternatives).

Lemma 6. *Let f be an Arrovian SWF, $a,b,c,d \in U$, $R \in \mathcal{D}^N$, and $\pi = (a,c)(b,d) \in \Pi(U)$. If $\pi(R|_{\{a,b\}}) = R|_{\{c,d\}}$, then $\phi(a,b) = \phi(c,d)$ where $\phi \equiv f(R)$.*

Proof. We first prove the case when all of a,b,c,d are distinct. Let $e \in U$ and consider a preference profile $\hat{R} \in \mathcal{D}^N$ such that $R|_{\{a,b,c,d\}} = \hat{R}|_{\{a,b,c,d\}}$ and $\hat{\phi}_i(x,e) = 1$ for all $x \in \{a,b,c,d\}$ and $i \in N$. Then, by Lemma 5, we can assume without loss of generality that $\hat{\phi}(x,e) = \lambda \in \mathbb{R}$ for all $x \in \{a,b,c,d\}$. Now consider a preference profile $\mathring{R} \in \mathcal{D}^N$ such that

$$\mathring{R}|_{\{a,b,c,d,e\}} \equiv (\underbrace{\begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 1 & 1 \\ -1 & -1 & 0 & 1 & 1 \\ -1 & -1 & -1 & 0 & 1 \\ -1 & -1 & -1 & -1 & 0 \end{pmatrix}, \dots,}_{N_{ab}} \underbrace{\begin{pmatrix} 0 & -1 & -1 & -1 & 1 \\ 1 & 0 & -1 & -1 & 1 \\ 1 & 1 & 0 & -1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ -1 & -1 & -1 & -1 & 0 \end{pmatrix}, \dots,}_{N_{ba}}$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ -1 & -1 & -1 & -1 & 0 \end{pmatrix}, \dots).$$

Note that $\hat{R}|_{\{a,b,e\}} = \mathring{R}|_{\{a,b,e\}}$ and $\hat{R}|_{\{c,d,e\}} = \mathring{R}|_{\{c,d,e\}}$ because $\pi(R|_{\{a,b\}}) = R|_{\{c,d\}}$ by assumption. Now, let $\hat{\phi} \equiv f(\hat{R})$ and $\mathring{\phi} \equiv f(\mathring{R})$. Since $\hat{R}|_{\{a,b,e\}} = \mathring{R}|_{\{a,b,e\}}$, we have $\hat{\phi}|_{\{a,b,e\}} \equiv \mathring{\phi}|_{\{a,b,e\}}$ by IIA. Moreover, $\hat{R}|_{\{c,d,e\}} = \mathring{R}|_{\{c,d,e\}}$ and IIA yield $\hat{\phi}|_{\{c,d,e\}} \equiv \mathring{\phi}|_{\{c,d,e\}}$. Lemma 5 implies that $\mathring{\phi}(x,e) = \lambda$ for some $\lambda \in \mathbb{R}$ for all $x \in \{a,b,c,d\}$. Thus, for some $\mu, \sigma \in \mathbb{R}$, $\mathring{\phi}$ takes the form

$$\mathring{\phi}|_{\{a,b,c,d,e\}} = \begin{pmatrix} 0 & \mu & & & \lambda \\ -\mu & 0 & & & \lambda \\ & & 0 & \sigma & \lambda \\ & & -\sigma & 0 & \lambda \\ -\lambda & -\lambda & -\lambda & -\lambda & 0 \end{pmatrix}.$$

Note that $\mathring{R}|_{\{a,b,c,d\}}$ only consists of one fixed preference relation, its inverse, and complete indifference. Hence, Pareto indifference and Lemma 1 imply that $\mathring{\phi}|_{\{a,b,c,d\}} = \alpha\phi_1|_{\{a,b,c,d\}}$ for some $\alpha \in \mathbb{R}$. Hence, we get that $\mu = \sigma$.

The cases when $a = c$ and $b = c$ follow from repeated application of the above case. All other cases are symmetric to one of the covered cases. □

Lemma 7. *Let f be an Arrovian SWF, $a, b, c, d \in U$, $R, \hat{R} \in \mathcal{D}^N$, $\phi \equiv f(R)$, and $\hat{\phi} \equiv f(\hat{R})$. If $R|_{\{a,b\}} = \hat{R}|_{\{a,b\}}$ and $R|_{\{c,d\}} = \hat{R}|_{\{c,d\}}$, then $\phi(a,b) = \alpha \cdot \hat{\phi}(a,b)$ and $\phi(c,d) = \alpha \cdot \hat{\phi}(c,d)$ for some $\alpha > 0$.*

Proof. Let $e \in U \backslash \{a,b,c,d\}$ and $R', \hat{R}' \in \mathcal{D}^N$ such that $R'|_{\{a,b,c,d\}} = R|_{\{a,b,c,d\}}$, $\hat{R}'|_{\{a,b,c,d\}} = \hat{R}|_{\{a,b,c,d\}}$, and $\phi'_i(x,e) = \hat{\phi}'_i(x,e) = 1$ for all $x \in \{a,b,c,d\}$ and $i \in N$. By $\phi' \equiv f(R')$ and $\hat{\phi}' \equiv f(\hat{R}')$ we denote the corresponding collective SSB functions. Since f satisfies IIA, we have that $\phi|_{\{a,b,c,d\}} \equiv \phi'|_{\{a,b,c,d\}}$ and $\hat{\phi}|_{\{a,b,c,d\}} \equiv \hat{\phi}'|_{\{a,b,c,d\}}$. Lemma 5 implies that without loss of generality, ϕ' and $\hat{\phi}'$ take the following form for some $\lambda, \mu, \hat{\mu}, \sigma, \hat{\sigma} \in \mathbb{R}$. Note that we can choose suitable representatives such that $\phi'(a,e) = \hat{\phi}'(a,e) = \lambda$.

$$\phi'|_{\{a,b,c,d,e\}} = \begin{pmatrix} 0 & \mu & & & \lambda \\ -\mu & 0 & & & \lambda \\ & & 0 & \sigma & \lambda \\ & & -\sigma & 0 & \lambda \\ -\lambda & -\lambda & -\lambda & -\lambda & 0 \end{pmatrix} \qquad \hat{\phi}'|_{\{a,b,c,d,e\}} = \begin{pmatrix} 0 & \hat{\mu} & & & \lambda \\ -\hat{\mu} & 0 & & & \lambda \\ & & 0 & \hat{\sigma} & \lambda \\ & & -\hat{\sigma} & 0 & \lambda \\ -\lambda & -\lambda & -\lambda & -\lambda & 0 \end{pmatrix}$$

Observe that $R'|_{\{a,b,e\}} = \hat{R}'|_{\{a,b,e\}}$ and $R'|_{\{c,d,e\}} = \hat{R}'|_{\{c,d,e\}}$ by construction. Since f satisfies IIA, we get that $\phi'|_{\{a,b,e\}} = \hat{\phi}'|_{\{a,b,e\}}$ and $\phi'|_{\{c,d,e\}} = \hat{\phi}'|_{\{c,d,e\}}$. In particular, this means that $\mu = \hat{\mu}$ and $\sigma = \hat{\sigma}$. Since $\phi|_{\{a,b,c,d\}} \equiv \phi'|_{\{a,b,c,d\}}$ and $\hat{\phi}|_{\{a,b,c,d\}} \equiv \hat{\phi}'|_{\{a,b,c,d\}}$, there is $\alpha \in \mathbb{R}$ as required. □

Lemma 7 shows that $\phi(a,b)$ only depends on N_{ab} and I_{ab} and not on a, b or R. Hence, there is a function $g \colon 2^N \times 2^N \to \mathbb{R}$ such that $g(N_{ab}, I_{ab}) = \phi(a,b)$ for all $a, b \in U$ and $R \in \mathcal{D}^N$ with $\phi \equiv f(R)$. We now leverage Pareto indifference to show that ϕ is a linear combination of the ϕ_i's. Hence, f is affine utilitarian.

Lemma 8. *Let f be an Arrovian SWF. Then, there are $w_1, \dots, w_n \in \mathbb{R}$ such that $f(R) \equiv \sum_{i \in N} w_i \phi_i$ for all $R \in \mathcal{D}^N$.*

Proof. For all $G \subseteq N$, let $w_G = {}^1\!/\!_2\,(g(N,\emptyset) + g(G,\emptyset))$. For convenience, we write w_i for $w_{\{i\}}$. Since $\phi(x,y) = g(N_{xy}, I_{xy})$ for all $x, y \in U$, it suffices to show that

$$g(N_{xy}, I_{xy}) = \sum_{i \in N} w_i \phi_i(x,y) = \sum_{i \in N_{xy}} w_i - \sum_{i \in N_{yx}} w_i, \tag{1}$$

for all $x, y \in U$. To this end, we will first show that $w_G + w_{\hat{G}} = w_{G \cup \hat{G}}$ for all $G, \hat{G} \subseteq N$ with $G \cap \hat{G} = \emptyset$. Let $G, \hat{G}$ as above, $a, b, c, x, y \in U$, and consider a following preference profile $R \in \mathcal{D}^N$ such that

$$R|_{\{a,b,c,x,y\}} \equiv (\underbrace{\begin{pmatrix} 0 & & & -1 & 1 \\ & 0 & & -1 & 1 \\ & & 0 & -1 & 1 \\ 1 & 1 & 1 & 0 & \\ -1 & -1 & -1 & & 0 \end{pmatrix}, \dots,}_{G} \underbrace{\begin{pmatrix} 0 & & & -1 & -1 \\ & 0 & & 1 & 1 \\ & & 0 & 1 & -1 \\ 1 & -1 & -1 & 0 & \\ 1 & -1 & 1 & & 0 \end{pmatrix}, \dots,}_{\hat{G}}$$

$$\begin{pmatrix} 0 & & & -1 & 1 \\ & 0 & & -1 & -1 \\ & & 0 & 1 & 1 \\ 1 & 1 & -1 & 0 & \\ -1 & 1 & -1 & & 0 \end{pmatrix}, \dots)$$

Let $\phi \equiv f(R)$. We have that, for $p = 1/2\, x + 1/2\, y$ and $q = 1/3\, a + 1/3\, b + 1/3\, c$, $\phi_i(p,q) = 0$ for all $i \in N$. Pareto indifference implies that $\phi(p,q) = 0$. Let $\mu = g(G, \emptyset)$, $\hat{\mu} = g(\hat{G}, \emptyset)$, and $\sigma = g(G \cup \hat{G}, \emptyset)$. By definition of w,

$$w_G + w_{\hat{G}} = w_{G \cup \hat{G}}$$

is equivalent to

$$(g(N,\emptyset) + g(G,\emptyset)) + (g(N,\emptyset) + g(\hat{G},\emptyset)) = g(N,\emptyset) + g(G \cup \hat{G},\emptyset).$$

Hence, we have to show that $\mu + \hat{\mu} + g(N,\emptyset) = \sigma$. By definition of g, we get that ϕ takes the following form.

$$\phi|_{\{a,b,c,x,y\}} = \begin{pmatrix} 0 & & & -g(N,\emptyset) & -\hat{\mu} \\ & 0 & & \hat{\mu} & \sigma \\ & & 0 & -\mu & -\hat{\mu} \\ g(N,\emptyset) & -\hat{\mu} & \mu & 0 & \\ \hat{\mu} & -\sigma & \hat{\mu} & & 0 \end{pmatrix}$$

From $\phi(p,q) = 0$, it follows that $1/6\, (\mu + \hat{\mu} + g(N,\emptyset) - \sigma) = 0$. This proves the desired relationship.

Now we can rewrite (1) as

$$g(N_{xy}, I_{xy}) = w(N_{xy}) - w(N_{yx}). \tag{2}$$

By definition of w, this is equivalent to

$$2g(N_{xy}, I_{xy}) = g(N_{xy},\emptyset) - g(N_{yx},\emptyset). \tag{3}$$

To prove (3), let $a, b, x, y \in U$ and consider a following preference profile $\hat{R} \in \mathcal{D}^N$ such that

$$\hat{R}|_{\{a,b,x,y\}} \equiv (\underbrace{\begin{pmatrix} 0 & & 1 & 1 \\ & 0 & -1 & -1 \\ -1 & 1 & 0 & \\ -1 & 1 & & 0 \end{pmatrix}, \dots,}_{G} \underbrace{\begin{pmatrix} 0 & & -1 & 1 \\ & 0 & -1 & 0 \\ 1 & 1 & 0 & \\ -1 & 0 & & 0 \end{pmatrix}, \dots,}_{\hat{G}} \begin{pmatrix} 0 & & -1 & -1 \\ & 0 & 1 & 1 \\ 1 & -1 & 0 & \\ 1 & -1 & & 0 \end{pmatrix}, \dots)$$

Let $\hat{\phi} \equiv f(\hat{R})$. Observe that, for $p = 1/3\, x + 2/3\, y$ and $q = 1/2\, a + 1/2\, b$, $\hat{\phi}_i(p, q) = 0$ for all $i \in N$. Pareto indifference implies that $\hat{\phi}(p, q) = 0$. With the same definitions as before and $\varepsilon = g(G, \hat{G})$, $\hat{\phi}$ takes the following form.

$$\hat{\phi}|_{\{a,b,x,y\}} \equiv \begin{pmatrix} 0 & & \mu & \sigma \\ & 0 & -\sigma & -\varepsilon \\ -\mu & \sigma & 0 & \\ -\sigma & \varepsilon & & 0 \end{pmatrix}$$

From $\hat{\phi}(p, q) = 0$, we get that $1/6\,(-\mu + \sigma - 2\sigma + 2\varepsilon) = 0$. Hence, $2\varepsilon = \mu + \sigma$. This is equivalent to

$$2g(G, \hat{G}) = g(G, \emptyset) + g(G \cup \hat{G}, \emptyset) = g(G, \emptyset) - g(N \setminus (G \cup \hat{G}), \emptyset),$$

where the last equality follows from skew-symmetry of $\hat{\phi}$ and the definition of g. This proves (3). □

Finally, the strict part of Pareto optimality implies that individual weights have to be strictly positive.

Theorem 2. *Let f be an Arrovian SWF. Then, there are $w_1, \dots, w_n \in \mathbb{R}_{>0}$ such that $f(R) \equiv \sum_{i \in N} w_i \phi_i$ for all $R \in \mathcal{D}^N$.*

Proof. From Lemma 8 we know that there are $w_1, \dots, w_n \in \mathbb{R}$ such that, for all $R \in \mathcal{R}^N$, $f(R) \equiv \sum_{i \in N} w_i \phi_i$. Assume for contradiction that $w_i \leq 0$ for some $i \in N$. Let G be the set of agents such that $w_i \leq 0$ and consider a preference profile $R \in \mathcal{D}^N$ with $a, b, c \in U$ such that

$$R|_{\{a,b,c\}} \equiv (\underbrace{\begin{pmatrix} 0 & & 1 \\ & 0 & 1 \\ -1 & -1 & 0 \end{pmatrix}, \dots,}_{G} \begin{pmatrix} 0 & & 1 \\ & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}, \dots).$$

Let $\phi \equiv f(R)$. Then, for $p = 1/2\, a + 1/2\, b$, we have that $\phi_i(p, c) > 0$ for all $i \in G$ and $\phi_i(p, c) = 0$ for all $i \in N \setminus G$. Pareto optimality of f implies that $\phi(p, c) > 0$. However, we have

$$\phi(p, c) = \alpha \left(\sum_{i \in G} w_i \phi_i(p, c) + \sum_{i \in N \setminus G} w_i \underbrace{\phi_i(p, c)}_{=0} \right) = \alpha \sum_{i \in G} \underbrace{w_i \phi_i(p, c)}_{\leq 0} \leq 0$$

for some $\alpha > 0$. This is a contradiction. □

D. Proof of Lemma 1

Before giving a proof of Lemma 1, we show four auxiliary statements about continuous and convex preference relations.

Lemma 9. *Let $\succ$ be a continuous and convex preference relation. Then $U(p)$ and $L(p)$ are open for all $p \in \Delta$.*

Proof. Let $p \in \Delta$. We start by showing that $I(p)$ is an affine subspace of Δ, i.e., $I(p) = \text{aff}(I(p)) \cap \Delta$.[19] To this end, let $q \in \text{aff}(I(p)) \cap \Delta$. Hence, there are $k \in \mathbb{N}$, $\lambda_i \in \mathbb{R}$ with $\sum_{i=1}^k \lambda_i = 1$, and $q_i \in I(p)$ such that $q = \sum_{i=1}^k \lambda_i q_i$. Equivalently,

$$r = \frac{1}{\sum_{i\colon \lambda_i \geq 0}} \left(q + \sum_{i\colon \lambda_i < 0} (-\lambda_i) q_i \right) = \frac{1}{\sum_{i\colon \lambda_i \geq 0}} \sum_{i\colon \lambda_i \geq 0} \lambda_i q_i.$$

Note that $1 + \sum_{i\colon \lambda_i<0} -\lambda_i = \sum_{i\colon \lambda_i \geq 0} \lambda_i$, since $\sum_{i=1}^k \lambda_i = 1$. Hence, $r \in \text{conv}(I(p))$ and thus, by convexity of $\succ$, $r \in I(p)$. If $q \in U(p)$, then, by convexity of $\succ$, $r \in U(p)$, which is a contradiction. Similarly, if $q \in L(p)$. Hence, $q \in I(p)$. This proves $I(p) = \text{aff}(I(p)) \cap \Delta$. Thus, as the intersection of two closed sets, $I(p)$ is closed.

Now assume for contradiction that $U(p)$ is not open, i.e., there is $q \in U(p)$ such that the ε-ball $B_\varepsilon(q)$ around q intersects with either $I(p)$ or $L(p)$ for every $\varepsilon > 0$. For $r \in B_\varepsilon(q) \cap L(p)$, by convexity of $\succ$ we have that, $\text{conv}(\{q, r\}) \cap I(p) \neq \emptyset$. Hence, $B_\varepsilon(q) \cap I(p) \neq \emptyset$ for all $\varepsilon > 0$. This implies that q is in the closure of $I(p)$, which contradicts the fact that $I(p)$ is closed. $\square$

Lemma 10. *Let $\succ$ be a continuous and convex preference relation. For all $p \in \Delta$, if $I(p)$ contains a non-empty open set, then $I(p) = \Delta$.*

Proof. Assume for contradiction that $I(p) \neq \Delta$ or, equivalently, $U(p) \cup L(p) \neq \emptyset$. Without loss of generality assume that $U(p) \neq \emptyset$. Let $q \in I(p)$ such that a neighborhood of q is contained in $I(p)$ and let $r \in U(p)$. Then convexity of $\succ$ implies that $\lambda q + (1-\lambda) r \in U(p)$ for all $\lambda \in (0, 1)$. This contradicts the assumption that a neighborhood of q is contained in $I(p)$. $\square$

The interior of a preference relation $\text{int}(\succ) = \{p \in \Delta \colon U(p) \neq \emptyset \text{ and } L(p) \neq \emptyset\}$ is the set of all outcomes with non-empty upper and lower contour sets.

Lemma 11. *Let $\succ$ be a continuous and convex preference relation. Then, for every $p \in \text{int}(\succ)$, $I(p) = \Delta \cap H$, where H is a $(|U|-1)$-dimensional hyperplane in $\mathbb{R}^U$. Moreover, $I(p)$ has dimension $|U| - 2$.*

[19] The affine hull $\text{aff}(X) = \{\sum_{i=1}^k \lambda_i x_i \colon k \in \mathbb{N}, \lambda_i \in \mathbb{R} \text{ with } \sum_{i=1}^k \lambda_i = 1, \text{ and } x_i \in X\}$ of $X \subseteq \mathbb{R}^U$ is the set of all affine combinations of elements in X.

Proof. Let $p \in \text{int}(\succ)$. Then, by Lemma 9, $U(p)$ and $L(p)$ are non-empty and open. Since $\succ$ is convex, $U(p)$ and $L(p)$ are convex. By the separating hyperplane theorem, there are $u \in \mathbb{R}^U$ and $s \in \mathbb{R}$ such that $H = \{v \in \mathbb{R}^U : u^T v = s\}$ strictly separates $U(p)$ and $L(p)$. Thus $\Delta \cap H \subseteq I(p)$. Since $U(p)$ and $L(p)$ are non-empty and H is strictly separating, H contains an interior point of Δ. Hence, $\Delta \cap H$ has dimension $|U| - 2$. If $I(p)$ has dimension $|U| - 1$, then, since $I(p)$ is convex, it contains an open subset of Δ. Lemma 10 implies that $I(p) = \Delta$. This contradicts $p \in \text{int}(\succ)$. ☐

Lemma 12. *Let $\succ$ be a continuous and convex preference relation. If $\succ \neq \emptyset$, then $\text{int}(\succ)$ is non-empty and open and $\text{cl}(\text{int}(\succ)) = \Delta$.*

Proof. First we show that $\text{int}(\succ) \neq \emptyset$. If $\succ \neq \emptyset$, there is $p \in \Delta$ such that $L(p) \neq \emptyset$. By Lemma 9, $L(p)$ is open. Let $q \in L(p)$. If $L(q) \neq \emptyset$, then $p \succ q \succ r$, i.e., $q \in \text{int}(\succ)$. Hence, consider the case that $L(q) = \emptyset$. If $L(p) \subseteq I(q)$, then $I(q)$ contains an open set and, by Lemma 10, $I(q) = \Delta$, which contradicts $q \in L(p)$. Thus, $L(p) \cap U(q) \neq \emptyset$. For $r \in L(p) \cap U(q)$, we have $p \succ r \succ q$, i.e., $r \in \text{int}(\succ)$.

To show that $\text{int}(\succ)$ is open, let $p \in \text{int}(\succ)$, $q \in U(p)$, and $r \in L(p)$. Then $p \in L(q) \cap U(r)$. Since, by Lemma 9, $L(q)$ and $U(r)$ are open and contain p, $L(q) \cap U(r)$ contains a neighborhood of p.

To show that $\text{cl}(\text{int}(\succ)) = \Delta$, let $p \in \max_\succ \Delta$. Let $O \subseteq \Delta$ be a neighborhood of p. Assume for contradiction that $O \cap \text{int}(\succ) = \emptyset$. If $O \cap (\min_\succ \Delta \setminus \max_\succ \Delta) \neq \emptyset$, let $q \in O \cap (\min_\succ \Delta \setminus \max_\succ \Delta)$. Since $q \notin \max_\succ \Delta$, it follows that $U(q) \neq \emptyset$. From Lemma 9 we know that $U(q)$ is open. Moreover, $q \in \text{cl}(U(q))$, since $\lambda q + (1 - \lambda)r \in U(q)$ for all $r \in U(q)$ and $\lambda > 0$. Hence, $O \cap U(q) \neq \emptyset$. As the intersection of open sets, $O \cap U(q)$ is open. Since $U(q) \cap \min_\succ \Delta = \emptyset$, the assumption that $O \cap \text{int}(\succ) = \emptyset$ implies that $O \cap U(q) \subseteq \max_\succ \Delta$. If $O \cap (\min_\succ \Delta \setminus \max_\succ \Delta) = \emptyset$, then, by assumption, $O \subseteq \max_\succ \Delta$. In any case, $\max_\succ \Delta$ contains an open set. Observe that, for all $p, q \in \max_\succ \Delta$, $q \in I(p)$. Now let $p \in O \subseteq \max_\succ \Delta$, where O is an open set and $q, r \in \Delta$ such that $r \in U(q)$. Then, by convexity, $(1 - \lambda)p + \lambda r \in U((1 - \lambda)p + \lambda q)$ for all $\lambda \in (0, 1)$. For small $\lambda > 0$, $(1 - \lambda)p + \lambda r$, $(1 - \lambda)p + \lambda q \in O \subseteq \max_\succ \Delta$, which contradicts $q \in I(p)$ for all $p, q \in \max_\succ \Delta$. Hence, for every $p \in \max_\succ \Delta$ and every neighborhood O of p, $O \cap \text{int}(\succ) \neq \emptyset$, i.e., $p \in \text{cl}(\text{int}(\succ))$. Similarly for $q \in \min_\succ \Delta$. Hence, $\text{cl}(\text{int}(\succ)) = \Delta$. ☐

We are now ready to prove Lemma 1.

Lemma 1. *Let $\succ, \hat{\succ}$ be continuous and convex preference relations. If $\sim \subseteq \hat{\sim}$, then $\hat{\succ} \in \{\succ, \succ^{-1}, \emptyset\}$.*

Proof. Let $p \in \Delta$. By assumption, we have $I(p) \subseteq \hat{I}(p)$. Moreover, Δ is the disjoint union of $I(p)$, $U(p)$, $L(p)$ and $\hat{I}(p)$, $\hat{U}(p)$, $\hat{L}(p)$, respectively. This implies that $\hat{U}(p) \cup \hat{L}(p) \subseteq U(p) \cup L(p)$. Assume for contradiction that $\hat{U}(p) \cap U(p) \neq \emptyset$ and $\hat{U}(p) \cap L(p) \neq \emptyset$. Let $q \in \hat{U}(p) \cap U(p)$ and $r \in \hat{U}(p) \cap L(p)$. Continuity of $\succ$ implies that $\text{conv}(\{q, r\}) \cap I(p) \neq \emptyset$. Convexity of $\hat{\succ}$ implies that $\text{conv}(\{q, r\}) \subseteq \hat{U}(p)$. Hence, $\emptyset \neq \text{conv}(\{q, r\}) \cap I(p) \subseteq \hat{U}(p)$, which contradicts $I(p) \subseteq \hat{I}(p)$. Hence, $\hat{U}(p) \subseteq U(p)$ or $\hat{U}(p) \subseteq L(p)$. Similarly, $\hat{L}(p) \subseteq L(p)$ or $\hat{L}(p) \subseteq U(p)$.

Now let $p \in \text{int}(\succ) \cap \text{int}(\hat{\succ})$. From Lemma 11, it follows that $I(p) = \Delta \cap H$ and $\hat{I}(p) = \Delta \cap \hat{H}$ for $(|U|-1)$-dimensional hyperplanes H and $\hat{H}$ through p. Moreover, $I(p)$ and $\hat{I}(p)$ have dimension $|U|-2$. Since $I(p) \subseteq \hat{I}(p)$, it follows that $I(p) = \hat{I}(p)$. Then, either $U(p) = \hat{U}(p)$ and $L(p) = \hat{L}(p)$ or $U(p) = \hat{L}(p)$ and $L(p) = \hat{U}(p)$. Let $\succ_p$ denote the restriction of $\succ$ to those comparisons involving p, i.e., $\succ_p = \succ \cap (\{p\} \times \Delta \cup \Delta \times \{p\})$. Thus, either $\succ_p = \hat{\succ}_p$ or $\succ_p = \hat{\succ}_p^{-1}$.

If $\hat{\succ} = \emptyset$, there is nothing left to show. Hence assume that $\hat{\succ} \neq \emptyset$. By assumption, this implies that $\succ \neq \emptyset$. From Lemma 12, it follows that $\text{int}(\succ) \cap \text{int}(\hat{\succ}) \neq \emptyset$. Let $p \in \text{int}(\succ) \cap \text{int}(\hat{\succ})$ and assume without loss of generality that $\succ_p = \hat{\succ}_p$. Let $q \in \text{int}(\succ) \cap \text{int}(\hat{\succ})$. If $q \in U(p) = \hat{U}(p)$, then $p \in L(q) \cap \hat{L}(q)$. Hence, $\succ_q = \hat{\succ}_q$. Similarly, if $q \in L(p)$. From Lemma 11, it follows that $I(p) \cup I(q) \neq \Delta$. Hence, $(U(p) \cup L(p)) \cap (\hat{U}(p) \cup \hat{L}(p))$ is non-empty and, by Lemma 9, open. By Lemma 12, $(U(p) \cup L(p)) \cap (\hat{U}(p) \cup \hat{L}(p)) \cap \text{int}(\succ)\text{int}(\hat{\succ})$ is non-empty and open. For $r \in (U(p) \cup L(p)) \cap (\hat{U}(p) \cup \hat{L}(p)) \cap \text{int}(\succ)\text{int}(\hat{\succ})$, it follows from two applications of what we have shown before that $\succ_r = \hat{\succ}_r$ and $\succ_q = \hat{\succ}_q$.

Now let $p \in \Delta \setminus (\text{int}(\succ) \cap \text{int}(\hat{\succ}))$. Assume for contradiction that $L(p) \setminus \hat{L}(p) \neq \emptyset$ and let $q \in L(p) \setminus \hat{L}(p)$. By Lemma 9, $L(p)$ is open. Hence, there is $\varepsilon > 0$ such that $B_\varepsilon(q) \subseteq L(p)$. If $B_\varepsilon(q) \cap \hat{L}(p) = \emptyset$, then $L(p) \setminus \hat{L}(p)$ contains an open set. If $B_\varepsilon(q) \cap \hat{L}(p) \neq \emptyset$, let $r \in B_\varepsilon(q) \cap \hat{L}(p)$. Since $B_\varepsilon(q) \cap \hat{L}(p)$ is the intersection of open sets, it is open. Hence, there is $\varepsilon' > 0$ such that $B_{\varepsilon'}(r) \subseteq B_\varepsilon(q) \cap \hat{L}(p)$. Let $\tau \colon B_\varepsilon(q) \to B_\varepsilon(q)$, $\tau(s) = q + (q - s)$ be the reflection with respect to q. Note that $q = {}^1\!/\!_2\,(s + \tau(s)) \in \text{conv}(\{s, \tau(s)\}$ for all $s \in B_\varepsilon(q)$. Hence, since convexity of $\hat{\succ}$ implies that $\hat{L}(p)$ is convex and $q \notin \hat{L}(p)$, $\tau(s) \in L(p) \setminus \hat{L}(p)$ for all $s \in B_{\varepsilon'}(r)$, i.e., $\tau(B_{\varepsilon'}(r)) \subseteq L(p) \setminus \hat{L}(p)$. In any case, there is an open set $O \subseteq L(p) \setminus \hat{L}(p)$. As the intersection of open sets, $O \cap \text{int}(\succ) \neq \emptyset$ is open. Since, by Lemma 12, $\text{cl}(\text{int}(\hat{\succ})) = \Delta$, it follows that $O \cap \text{int}(\succ) \cap \text{int}(\hat{\succ}) \neq \emptyset$. Thus, there is $q \in \text{int}(\succ) \cap \text{int}(\hat{\succ})$ such that $q \in L(p)$ but $q \notin \hat{L}(p)$. From before we know that $\succ_r = \hat{\succ}_r$ for all $r \in \text{int}(\succ) \cap \text{int}(\hat{\succ})$, which is a contradiction. Hence, $\hat{L}(p) = L(p)$. Similarly, we get $\hat{U}(p) = U(p)$. In summary, $\hat{L}(p) = L(p)$, $\hat{U}(p) = U(p)$, and $I(p) \subseteq \hat{I}(p)$, which implies that $\succ_p = \hat{\succ}_p$. □

Lemma 1 does not hold if convexity is weakened to the assumption that $U(p)$, $L(p)$, and $I(p)$ need to be convex for all $p \in \Delta$. To see this, consider the following preference relations on the closed interval $[0,1]$ (equipped with the standard topology). Let $\succsim$ be the greater or equal relation and $\hat{\succsim}$ be defined such that $x \mathrel{\hat{\succ}} y$ if $x \in ({}^3\!/\!_4, 1]$ and $y \in [0, {}^1\!/\!_4)$ and $x \mathrel{\hat{\sim}} y$ otherwise. Both, $\succsim$ and $\hat{\succsim}$ are continuous and convex according to the weaker convexity assumption defined above. For $\succsim$ this is clear. To see this for $\hat{\succsim}$, observe that for all $x \in [0,1]$, either $I(x) = [0, {}^3\!/\!_4]$ and $U(x) = ({}^3\!/\!_4, 1]$ or $I(x) = [0,1]$ or $L(x) = [0, {}^1\!/\!_4)$ and $I(x) = [{}^1\!/\!_4, 1]$. In all cases, $U(x)$ and $L(x)$ are open and $U(x)$, $L(x)$, and $I(x)$ are convex.