Belief-weighted Nash Aggregation of Savage Preferences

Yves Sprumont*

September 23, 2018

Abstract

The *belief-weighted Nash social welfare functions* are methods for aggregating Savage preferences defined over a set of acts. Each such method works as follows. Fix a 0-normalized subjective expected utility representation of every possible preference and assign a vector of individual weights to each profile of beliefs. To compute the social preference at a given preference profile, rank the acts according to the weighted product of the individual 0-normalized subjective expected utilities they yield, where the weights are those associated with the belief profile generated by the preference profile. We show that these social welfare functions are characterized by the weak Pareto principle, a continuity axiom, and the following informational robustness property: the social ranking of two acts is unaffected by the addition of any outcome that every individual deems at least as good as the one she originally found worst. This makes the belief-weighted Nash social welfare functions appealing in contexts where the *best* relevant outcome for an individual is difficult to identify.
Keywords: preference aggregation, uncertainty, subjective expected utility, Nash product.

JEL classification numbers: D63, D71.

1. Introduction

1.1. Context and related work

This note reconsiders the problem of aggregating preferences obeying the axioms of Savage's theory of choice under uncertainty. In that theory, uncertain prospects are modeled as acts, namely, mappings from states of nature to outcomes, and an individual's preference is summarized by her subjective assessment of the likelihood of the possible events and the

*Département de Sciences Économiques and CIREQ, Université de Montréal, C.P. 6128, succursale centre-ville, Montréal QC, H3C 3J7, Canada (yves.sprumont@umontreal.ca). I thank M. Amarante and S. Horan for many discussions on the topic of this note. I also thank two referees, an associate editor, W. Bossert, C. Chambers, M. Kaneko, and P. Mongin for useful comments, and the FRQSC for financial support.

utility she attaches to the conceivable outcomes: she compares acts according to their subjective expected utility. Aggregating such "Savage" preferences is notoriously problematic. Mongin (1995) shows that any rule which transforms a collection of Savage preferences into a Savage social preference and respects the Pareto (indifference) criterion –that is, deems two acts equally good when all individuals do– must be radically uncompromising: at most profiles, the social preference coincides with the preference of one of the individuals.[1] In reaction to this incompatibility result, one may follow the "Savage approach" –that is, focus on aggregation rules that violate the Pareto criterion but guarantee a Savage social preference– or the "Pareto approach" –focus on rules that satisfy the Pareto criterion but need not yield a Savage social preference.[2]

The Savage approach is motivated by Mongin's (1997) criticism of the Pareto criterion. He argues that if two individuals agree on the comparison of two acts only because the differences between their subjective beliefs compensate the differences between their utilities, society should not be bound by the "spurious unanimity" of their preferences. Building on that criticism, Gilboa, Samet and Schmeidler (2004) suggest that the Pareto criterion should be respected only when the individuals agree on the probabilities of the events relevant to the uncertain prospects they compare. This weakened Pareto criterion implies that if society's preference is Savage, its utility for the outcomes is a weighted sum of its members' utilities, and its belief a weighted sum of their beliefs. Gilboa, Samuelson and Schmeidler (2014) study a different weakening of the Pareto criterion.

Relaxing the Pareto criterion makes sense if the individuals' probability measures do indeed represent their assessments of the likelihood of the events: when such assessments differ, a least one individual must be mistaken and society should not be compelled to respect the unanimous preferences of its members because they cannot all be well informed. In Savage's theory, however, a subjective probability measure is just an abstract system of weights. These weights may reflect in part the individual's assessment of the likelihood of the events, but they may reflect other subjective considerations as well.[3] This point is made, for instance, by Duffie (2014). In such contexts, no subjective probability measure can be wrong, and dropping the Pareto criterion is dangerous. This is the motivation for the Pareto approach.[4]

[1] Hylland and Zeckhauser (1979) establish a similar result in a multi-profile context. Chambers and Hayashi (2006) show that, at some profiles, Paretian aggregation is impossible even if society's preference is only required to satisfy Savage's P3 or P4 axiom.

[2] The Savage approach is often called the *ex-post approach* as it yields ex-post (i.e., conditional) social preferences satisfying the Pareto principle with respect to the individual ex-post preferences. Likewise, the Pareto approach is often called the *ex-ante approach* as it imposes the Pareto principle with respect to the individual ex-ante preferences.

[3] A related difficulty is that a given preference satisfying Savage's axioms admits, on top of Savage's representation, multiple *state-dependent* expected utility representations. If utility is indeed state-dependent, then the probability measure from Savage's theorem (derived under the wrong assumption that utility is state-independent) does not correctly reflect the individual's beliefs. Savage and Aumann (1987) discuss the example of a man whose wife is gravely ill. He may find the event that she dies very likely, yet attach a low weight to it –reflecting the fact that he does not enjoy life without her. See Baccelli (2017) for a comprehensive and recent discussion.

[4] We have nothing to add to the debate about the Pareto criterion under uncertainty. There is evidence

Mongin (1997) shows that the Pareto criterion can be respected if society's preference is allowed to be of the *state-dependent* subjective expected utility type. When that is the case, Chambers and Hayashi (2006) prove that Pareto indifference implies that the social utility function over the set of acts is a weighted sum of the utility functions of its members.

To the best of our knowledge, and regardless of their relative merits, neither the Savage nor the Pareto approach have so far offered a complete resolution of the problem of aggregating Savage preferences. Results such as those of Gilboa, Samet and Schmeidler (2004) and Chambers and Hayashi (2006) do not tell us (i) how the individual utilities should be calibrated and (ii) how this calibration should depend on the preference profile.[5] Therefore, they do not define a *social welfare function* (or SWF, for short), namely, a mapping assigning a social preference to *every* profile of individual Savage preferences. This is the criticism that Dhillon and Mertens (1999) already formulated against Harsanyi's (1955) theorem on the utilitarian aggregation of von Neumann-Morgenstern preferences over lotteries: since no restriction is imposed on how utilities are calibrated, *"the 'individual utility functions' become arbitrarily complex functions of the preferences of all other individuals, making the sum-formula basically meaningless."*.

1.2. Our contribution

As a partial solution to that problem, we define and axiomatize the class of *belief-weighted Nash SWFs*. The simplest example of such a SWF uses fixed equal weights. For each Savage preference over acts, choose a 0-normalized subjective expected utility representation –one where the utility of the worst outcome for that preference is zero.[6] At any preference profile, rank the acts according to the product of the individual 0-normalized subjective expected utilities they yield. There are infinitely many possible 0-normalizations for each preference but the social ranking does not change with the chosen normalizations because all 0-normalized utilities associated with a given preference are a positive multiple of each other and society's utility is their product.[7]

More generally, under a *belief-weighted* Nash SWF, an act is evaluated according to a weighted product of the individual 0-normalized subjective expected utilities it yields. The weights depend on the profile of *beliefs* of the individuals, but not on their utilities

that individuals differ in their assessments of the likelihood of many events, and we agree that the Pareto criterion is problematic when such differences exist. At the same time, we believe that a satisfactory weakening of it requires a theoretical model of behavior where the "likelihood assessment" component of an individual probability measure can be formally disentangled from its "residual" component. A recent paper that addresses this problem is Mongin and Pivato (2016).

[5] This remains true even if the weights in these representation theorems have somehow been determined.

[6] In the formal analysis below, we do not assume that a worst outcome exists. Rather, we suppose that an individual's preference is representable by a bounded utility function, and the 0-normalization means that the infimum of such a function is zero.

[7] Note that the worst outcome need not be the same for all individuals, and the measure of an individual's welfare is relative to the outcome *she* finds worse. This is in contrast with Nash's (1950) analysis of the bargaining problem, where the disagreement utility vector is generated by an outcome that is the same for all individuals.

for the outcomes.

Our axiomatization of this class of SWFs falls squarely in the Pareto approach. We impose the weak Pareto principle but no restriction on the social preference beyond the condition that it must be an ordering. Of course, there are many Paretian SWFs. In order to understand what makes the belief-weighted Nash SWFs special, it is instructive to first take a look at another natural solution: relative utilitarianism. Adapted to the context of uncertainty, relative utilitarianism compares two acts according to the sums of the $(0,1)$-normalized subjective expected utilities they yield –the latter normalization consisting in assigning utility 0 to any outcome that an individual deems worst and 1 to any outcome she finds best.[8] Observe that, in contrast to the belief-weighted Nash SWFs, a double normalization is required to obtain a well-defined SWF.

Under relative utilitarianism, the recommended social ranking of two acts may change with the set of outcomes that are considered relevant. In particular, it may be affected by the addition of an outcome that an individual deems worse than the one she initially found worst, *or better than the one she initially found best.*

In many applications, identifying the worst relevant outcome for each individual may be relatively easy, but determining the best one is difficult. As an illustration, consider the problem of developing medical treatment against two diseases, A and B. Let x_d denote the quality of the treatment developed against disease d : say that $x_d = 0$ if no treatment exists, $x_d = \frac{1}{2}$ if a good treatment is made available, and $x_d = 1$ if the treatment is excellent (these numbers are a convenient way of indexing the possibilities but have no meaning –we could use $x_d = \alpha, \beta, \gamma$ instead). The relevant outcomes are all the pairs $x = (x_A, x_B)$ in the set

$$X = \left\{0, \frac{1}{2}, 1\right\} \times \left\{0, \frac{1}{2}, 1\right\}.$$

These outcomes are uncertain and health policies may be regarded as acts mapping states of nature into X. The precise specification of the set of relevant states of nature is irrelevant for the point we want to make. There are two individuals with Savage preferences over the set of acts. The specification of their beliefs is also unimportant. Individual 1 suffers from disease A; her utility for the outcomes is given by the function $u_1(x) = x_A$ for all $x = (x_A, x_B) \in X$. Individual 2 suffers from disease B and her utility for the outcomes is $u_2(x) = x_B$ for all $x \in X$. Observe that, given X, the functions u_1, u_2 are $(0,1)$-normalized: $\inf_X u_i = 0$ and $\sup_X u_i = 1$ for $i = 1, 2$. Relative utilitarianism deems the (constant acts producing in all states of nature the) outcomes $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$ equally good[9] because both

[8] Relative utilitarianism was originally proposed in the context of risk. In that framework, it evaluates a lottery according to the sum of the $(0,1)$-normalized von Neumann-Morgenstern utilities it generates. The criterion has received several axiomatizations: see Dhillon (1998), Karni (1998), Dhillon and Mertens (1999), Segal (2000), and Börgers and Choo (2017a, 2017b). All these papers assume that society's preference over lotteries is of the von Neumann-Morgenstern type. In the context of uncertainty, and as a corollary to Mongin's (1995) theorem, relative utilitarianism does not always produce a Savage social preference over acts. It is therefore completely unclear how the axiomatizations proposed in the context of risk could be adapted to uncertainty.

[9] As usual, we identify a constant act with the outcome it yields in all states.

generate a sum of $(0,1)$-normalized (subjective expected) utilities equal to $\frac{1}{2}$.

Suppose now that, in fact, an excellent treatment cannot possibly be developed against B. The set of relevant outcomes then becomes

$$Y = \left\{0, \frac{1}{2}, 1\right\} \times \left\{0, \frac{1}{2}\right\}.$$

Given Y, the individual $(0,1)$-normalized utility functions over the set of outcomes are now $v_1(x) = u_1(x) = x_A$ and $v_2(x) = 2u_1(x) = 2x_B$. Relative utilitarianism deems $(0, \frac{1}{2})$ preferable to $(\frac{1}{2}, 0)$. Thus, in order to decide whether a good treatment against A (and no treatment against B) is preferable to a good treatment against B (and no treatment against A), society needs to know whether an excellent treatment against B (and A) is possible or not. There need not be anything morally wrong with this view, but it may be difficult to implement in practice. In many contexts, such as the one above, the best relevant outcome for each individual is hard to determine.

In the example above, the *Nash SWF* (with, say, fixed equal weights) is more appealing. Because u_1, u_2 are 0-normalized for both X and Y (that is, $\inf_X u_i = \inf_Y u_i = 0$ for $i = 1, 2$), the Nash SWF deems the outcomes $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$ equally good independently of whether the set of relevant outcomes is X or Y. Of course, determining the social preference still requires a correct specification of the worst relevant outcome for each individual. In many cases, this may not be an impossible task.

The example illustrates a general property of the Nash ranking: it is unaffected by the addition of any outcome that all individuals find at least as good as the one they initially found worst –*even if such a new outcome is better than the one they initially found best.* As discussed above, this property is compelling when the best outcome for an individual cannot be determined with confidence, an arguably frequent case. In this note, we show that the belief-weighted Nash SWFs are the only weakly Paretian SWFs satisfying this "Independence of Harmless Expansions" property and a continuity condition to be described below.

1.3. More connections with the literature

Kaneko and Nakamura (1979) axiomatize the Nash SWF for aggregating von Neumann-Morgenstern preferences over lotteries. As we have seen in Sub-section 1.1, aggregating Savage preferences over acts is a quite different exercise. Indeed, the (non-degenerate) *belief-weighted* SWFs have no counterpart in the lottery framework. There are two further major differences between Kaneko and Nakamura's work and ours. The first and most important one is conceptual: their analysis is restricted to problems where all individuals agree on what the worst relevant outcome is: a lottery is then evaluated according to the product of the von Neumann-Morgenstern utility gains it generates with respect to this *common* worst outcome. The scope of applicability of the SWFs we define is much broader: they allow society to rank the relevant acts for *any* profile of Savage preferences. The second remaining difference is more technical, though also important. Kaneko and

Nakamura use an independence axiom embodying an assumption of neutrality which our Independence of Harmless Expansions axiom completely dispenses with.

Another related paper is West (1984). The author considers the problem of ranking social *gambles*, namely, acts that return a positive amount of money to each individual if a given event occurs, and zero to everyone otherwise. In this very special context, it turns out that (the suitable version of) Pareto indifference[10] is compatible with the requirement that society has well-defined beliefs and tastes, and the author shows that the latter requirement essentially forces a multiplicative aggregation of the individual utilities and beliefs. This argument cannot be used in our framework because of Chambers and Hayashi's (2006) impossibility results.

2. Framework

Let S be the set of possible *states (of nature)*. Subsets of S are called *events*. Let $\mathbb{X}$ be an uncountable set of *conceivable (social) outcomes*, and let $\mathcal{X}$ be the set of finite or countably infinite subsets of $\mathbb{X}$ containing at least two elements. For each $X \in \mathcal{X}$, let $A(X)$ be the set of functions from S to X. Elements of X are called *relevant outcomes* and elements of $A(X)$ *relevant acts.* If $x \in X$, we abuse notation and also use x to denote the constant act assigning outcome x to every state.

For any $X \in \mathcal{X}$, a *preference* over $A(X)$ is an ordering $R \subseteq A(X) \times A(X)$. We call R *Savage* if there exist a non-constant, bounded function $u : X \to \mathbb{R}$ and a finitely additive, non-atomic probability measure p on 2^S –henceforth called a *belief*– such that

$$aRb \Leftrightarrow \int_S (u \circ a)dp \geq \int_S (u \circ b)dp$$

for all $a, b \in A(X)$. We let $\mathcal{P}$ denote the set of all beliefs. The function $U(., u, p) : A(X) \to \mathbb{R}$ defined by

$$U(a, u, p) := \int_S (u \circ a)dp \text{ for all } a \in A(X)$$

is a *Savage representation* of R. We denote by $\mathcal{U}(X, R)$ the set of such representations. If $U(., u, p)$ and $U(., v, q)$ are two Savage representations of R, then $p = q$ and $v = \alpha u + \beta$ for some positive real number α and some real number β. We let $p^*(R)$ denote the unique belief $p \in \mathcal{P}$ such that $p = p^*(R)$ for every Savage representation $U(., u, p)$ of R. Let $\mathcal{R}(X)$ denote the set of all preferences over $A(X)$ and $\mathcal{R}^*(X)$ the subset of Savage preferences. Write $\mathcal{R} = \cup_{X \in \mathcal{X}} \mathcal{R}(X)$ and $\mathcal{R}^* = \cup_{X \in \mathcal{X}} \mathcal{R}^*(X)$.

Let $N = \{1, ..., n\}$ be a finite set of individuals. A *(social choice) problem* is a list (X, R_N) where $X \in \mathcal{X}$ and $R_N = (R_1, ..., R_n) \in \mathcal{R}^*(X)^N$. We simply call R_N a *preference profile* (over $A(X)$) –but keep in mind that $R_1, ..., R_n$ are *Savage* preferences. The set (or

[10] If each individual i is indifferent between the gamble $(x_i, 0)$ and the sure monetary payoff y_i, then society is indifferent between the social gamble $((x_1, ..., x_n), (0, ..., 0))$ and the vector of monetary payoffs $(y_1, ..., y_n)$.

domain) of all problems is denoted by $\mathcal{D}$. A *social welfare function (or SWF)* is a mapping $\mathbf{R} : \mathcal{D} \to \mathcal{R}$ such that $\mathbf{R}(X, R_N) \in \mathcal{R}(X)$ for every $(X, R_N) \in \mathcal{D}$.

A few comments are in order about the setting just described.

(1) In the spirit of Arrow (1963), a SWF is a completely ordinal object that aggregates preference *orderings*: no utility information is available. We interpret $\mathbf{R}(X, R_N)$ as *society's preference* over $A(X)$ when individual preferences are given by the profile R_N.

In contrast to Mongin (1995), the preference profile R_N is variable. This variable-profile approach has a long tradition in social choice theory where recommending a SWF is interpreted as designing a "constitution" –a fully specified procedure for solving not just one but any possible preference aggregation problem. The underlying view is that, in order to avoid arbitrariness, the social decision maker must first commit to such a formal procedure before asking individuals to report their preferences.

Clearly, the variable-profile approach cannot avoid the incompatibility between the Pareto principle and collective Savage rationality. But it allows one to formulate collective rationality requirements that vary with the profile that society's preference summarizes. The basic idea is that society can afford to be more rational when its members agree than when they do not. More on this in Subsection 5.3.

(2) In contrast to the standard Arrovian formulation, the set over which society's preference is constructed is allowed to vary. When the set of relevant outcomes expands, society's preference over the originally relevant acts is *a priori* allowed to change: if $X \subseteq X'$ and the preference profile R'_N over $A(X')$ coincides over $A(X)$ with the profile R_N, $\mathbf{R}(X', R'_N)$ need not coincide with $\mathbf{R}(X, R_N)$ on $A(X)$. The axiom of Independence of Harmless Expansions discussed in the Introduction and formally defined in the next section will restrict the extent to which society's preference is allowed to change.

Note also that society's preference is constructed on sets of acts with an at most countable range of outcomes X whereas the set of *conceivable* outcomes $\mathbb{X}$ is uncountable. That relevant alternatives always form a "small" subset of all the conceivable ones is perhaps not a bad assumption.[11]

(3) Society's preference over $A(X)$ may only depend upon individual preferences *over that set*. This is a natural restriction because $\mathbb{X}$ is a large unstructured set and individual preferences over acts whose outcomes belong to $\mathbb{X}$ may therefore be difficult to elucidate. Moreover, since no structure is imposed on $\mathbb{X}$, there is no natural reference point outside $A(X)$ which could help define the social preference on $A(X)$.

(4) Individual preferences are of the Savage type but society's preference need not be. The set of possible SWFs is therefore quite large, and axioms such as the Pareto criterion should not be expected to have much bite.

[11] The assumption that $\mathbb{X}$ is uncountable is used below in the proof of the neutrality lemma and in steps 1 and 2.3 of the proof of our main result. The assumption is stronger than necessary and can be replaced by the condition that $\mathcal{X}$ is a sigma-ideal of subsets of $\mathbb{X}$.

3. Theorem

For any $X \in \mathcal{X}$ and $R \in \mathcal{R}^*(X)$, a Savage representation $U(.,u,p)$ of R is 0-*normalized* if $\inf_X u = 0$; it is $(0,1)$-normalized if in addition $\sup_X u = 1$. We denote by $\mathcal{U}_0(X,R)$ the set of 0-normalized Savage representations of R and by $U^*(.,X,R)$ the unique $(0,1)$-normalized Savage representation of R. If $(X,R_N) \in \mathcal{D}$ and $U_i \in \mathcal{U}(X,R_i)$ for each $i \in N$, define $U_N : A(X) \to \mathbb{R}^N_+$ by $U_N(a) = (U_1(a),...,U_n(a))$ for all $a \in A(X)$. With a slight abuse of notation, write $p^*(R_N) := (p^*(R_1),...,p^*(R_n))$, $\mathcal{U}(X,R_N) := \prod_{i\in N} \mathcal{U}(X,R_i)$, and $\mathcal{U}_0(X,R_N) := \prod_{i\in N} \mathcal{U}_0(X,R_i)$. Let $\Delta_N = \left\{\gamma \in [0,1]^N \mid \sum_{i\in N} \gamma_i = 1\right\}$ denote the simplex in $\mathbb{R}^N$.

A SWF $\mathbf{R}$ is a *belief-weighted Nash SWF* if there is a function $\gamma : \mathcal{P}^N \to \Delta_N$ such that, for all $(X,R_N) \in \mathcal{D}$ and all $a,b \in A(X)$,

$$a\mathbf{R}(X,R_N)b \Leftrightarrow \left[\prod_{i\in N} U_i(a)^{\gamma_i(p^*(R_N))} \geq \prod_{i\in N} U_i(b)^{\gamma_i(p^*(R_N))} \text{ for all } U_N \in \mathcal{U}_0(X,R_N)\right]. \quad (3.1)$$

We call γ the *weight function* associated with $\mathbf{R}$. Since for every $U_i \in \mathcal{U}_0(X,R_i)$ there exists a positive real number α_i such that $U_i = \alpha_i U^*(.,X,R_i)$, (3.1) is equivalent to

$$a\mathbf{R}(X,R_N)b \Leftrightarrow \prod_{i\in N} U^*(a,X,R_i)^{\gamma_i(p^*(R_N))} \geq \prod_{i\in N} U^*(b,X,R_i)^{\gamma_i(p^*(R_N))}.$$

Some further notation and terminology is needed to state our axiomatic characterization of the belief-weighted Nash SWFs. We let $\mathbb{N} := \{1,2,...\}$ denote the set of positive integers. The symbols P_i and I_i denote the strict preference and indifference relations associated with the individual preference R_i, and $\mathbf{P}(X,R_N)$ and $\mathbf{I}(X,R_N)$ are the strict social preference and indifference relations associated with $\mathbf{R}(X,R_N)$. If $(X,R_N),(X',R'_N) \in \mathcal{D}$ and $X \subseteq X'$, we say that R'_N *coincides with* R_N *on* $A(X)$ if $R'_i \cap (A(X)\times A(X)) = R_i$ for all $i \in N$. Similarly, $\mathbf{R}(X',R'_N)$ coincides with $\mathbf{R}(X,R_N)$ on $A(X)$ if $\mathbf{R}(X',R'_N) \cap (A(X)\times A(X)) = \mathbf{R}(X,R_N)$. Finally, if $(X,R_N) \in \mathcal{D}$, $a \in A(X)$, and (a^t) is a sequence of acts in $A(X)$, we say that (a^t) *converges to* a *with respect to* R_N if the sequence $(U_N(a^t))$ converges to $U_N(a)$ for all $U_N \in \mathcal{U}(X,R_N)$.

We now state the axioms used in our characterization theorem. The first needs no introduction; part (ii) of this axiom is known as *Pareto Indifference*.

Weak Pareto Principle.[12] For all $(X,R_N) \in \mathcal{D}$ and all $a,b \in A(X)$, (i) if aP_ib for all $i \in N$, then $a\mathbf{P}(X,R_N)b$, and (ii) if aI_ib for all $i \in N$, then $a\mathbf{I}(X,R_N)b$.

The second axiom plays a central role in our characterization.

[12] The Weak Pareto Principle is often called the *ex-ante Pareto axiom* as it requires unanimity with respect to the individuals' ex-ante preferences (over acts). It is stronger than the *ex-post Pareto axiom* which only requires unanimity with respect to the individuals' ex-post preferences (over outcomes).

Independence of Harmless Expansions. For all $(X, R_N), (X', R'_N) \in \mathcal{D}$, if (i) $X \subseteq X'$, (ii) R'_N coincides with R_N on $A(X)$, and (iii) for all $x' \in X'$ and $i \in N$ there exists some $x_i \in X$ such that $x' R'_i x_i$, then $\mathbf{R}(X', R'_N)$ coincides with $\mathbf{R}(X, R_N)$ on $A(X)$.

The motivation for this axiom was already presented in the Introduction. Although the SWF $\mathbf{R}$ is requested to produce a social ordering for every problem in $\mathcal{D}$, the social decision maker may in fact be uncertain about the correct specification of the problem she is facing. The axiom assumes that this indeed occurs when two problems (X, R_N), (X', R'_N) are related by conditions (i), (ii), and (iii). Since the decision maker is unable to assert whether the correct formulation is (X, R_N) or (X', R'_N), the ordering $\mathbf{R}(X', R'_N)$ should coincide with $\mathbf{R}(X, R_N)$ on $A(X)$.

Independence of Harmless Expansions is an independence condition very much akin to Arrow's Independence of Irrelevant alternatives. Indeed, Arrow's axiom in our setting would correspond to the statement obtained by dropping proviso (iii) in our axiom.

Independence of Harmless Expansions is also related to Dhillon and Mertens' (1999) Independence of Redundant Alternatives. In our framework, the latter axiom would require that if (i) $X \subseteq X'$, (ii) R'_N coincides with R_N on $A(X)$, and (iii') for all $x' \in X'$ there exists some $x \in X$ such that $x' I'_i x$ for all $i \in N$, then $\mathbf{R}(X', R'_N)$ coincides with $\mathbf{R}(X, R_N)$ on $A(X)$. This is a weaker axiom than Independence of Harmless Expansions because condition (iii') is more restrictive than (iii) in two respects: it imposes not only that, for each individual, each outcome x' in X' be at least as good as the worst outcome in X, but that there be a *common* outcome in X that all individuals deem *equivalent* to x'.

Our axiom is also related to Kaneko and Nakamura's (1979) Independence of Irrelevant Alternatives with Neutral Property. The latter is a stronger and rather complicated axiom that mixes the same independence condition as ours with the requirement that the names of the outcomes should not affect the social ordering.

Finally, it may be worth pointing out the apparent similarity between our axiom and Nash's Independence of Irrelevant Alternatives. This similarity is misleading because the objects on which the conditions are formulated are very different. As a matter of fact, the primary role of Nash's independence axiom is to establish that the solution to the bargaining problem must be maximizing some underlying ordering defined over utility space; the fact that this ordering is the product ordering follows from Nash's scale invariance axiom, not from his independence axiom. By contrast, Independence of Harmless Expansions is the axiom responsible for the product form in our characterization.

Our third axiom requires a form of continuity of the social preference at every given preference profile. It does not impose any restriction across profiles.

Continuity. For all $(X, R_N) \in \mathcal{D}$, all $a, b \in A(X)$, and every sequence (a^t) in $A(X)$ converging to a with respect to R_N, (i) if $a^t \mathbf{R}(X, R_N) b$ for all $t \in \mathbb{N}$, then $a \mathbf{R}(X, R_N) b$, and (ii) if $b \mathbf{R}(X, R_N) a^t$ for all $t \in \mathbb{N}$, then $b \mathbf{R}(X, R_N) a$.

Theorem. *A SWF satisfies the Weak Pareto Principle, Independence of Harmless Expansions, and Continuity if and only if it is a belief-weighted Nash SWF.*

We emphasize that our axioms perform three tasks: (i) they force us to use 0-normalized

Savage representations of the individual preferences *at all preference profiles*[13], (ii) they imply that social welfare is a *weighted product* of these individual 0-normalized Savage utilities, and (iii) they imply that the weights attached to the individual Savage utilities can only depend on the profile of *beliefs*. The only (but important) remaining indeterminacy lies in the choice of the function that computes the weights for every profile of beliefs. This indeterminacy can be somewhat reduced by imposing further axioms: see Section 5.3 for a discussion.

4. Proof

In order to prove our theorem, we begin with a lemma showing that Pareto Indifference and Independence of Harmless Expansions imply a strong form of outcome neutrality.[14] Let $\Pi(\mathbb{X})$ denote the set of bijections from $\mathbb{X}$ into itself. If $(X, R_N) \in \mathcal{D}$, $\pi \in \Pi(\mathbb{X})$, $a \in A(X)$, and $R_N \in \mathcal{R}^*(X)^N$, let $a^\pi \in A(\pi(X))$ be the act given by $a^\pi(s) = \pi(a(s))$ for all $s \in S$, and denote by R_N^π the preference profile on $A(\pi(X))$ given by $a^\pi R_i^\pi b^\pi \Leftrightarrow aR_ib$ for all $i \in N$ and all $a, b \in A(X)$.

Outcome Neutrality. For all $(X, R_N) \in \mathcal{D}$, $a, b \in A(X)$ and $\pi \in \Pi(\mathbb{X})$, $a\mathbf{R}(X, R_N)b \Leftrightarrow a^\pi\mathbf{R}(\pi(X),\, R_N^\pi)b^\pi$.

Lemma. *If a SWF satisfies Pareto Indifference and Independence of Harmless Expansions, then it satisfies Outcome Neutrality.*

Proof. Let $\mathbf{R}$ satisfy Pareto Indifference and Independence of Harmless Expansions. Let $(X, R_N) \in \mathcal{D}$, $a, b \in A(X)$ and $\pi \in \Pi(\mathbb{X})$. We prove that $a\mathbf{R}(X, R_N)b \Rightarrow a^\pi\mathbf{R}(\pi(X), R_N^\pi)b^\pi$. The converse implication follows immediately since $a = (a^\pi)^{\pi^{-1}}$, $b = (b^\pi)^{\pi^{-1}}$, $X = \pi^{-1}(\pi(X))$, and $R_N = (R_N^\pi)^{\pi^{-1}}$. Let us thus assume that

$$a\mathbf{R}(X, R_N)b. \tag{4.1}$$

Step 1. $a^\pi\mathbf{R}(\pi(X), R_N^\pi)b^\pi$ *if* $\pi(X) \cap X = \emptyset$.

Let $\overline{X} = X \cup \pi(X)$. For each $i \in N$, let $\overline{R}_i$ be the Savage preference over $A(\overline{X})$ which coincides with R_i on $A(X)$ and is such that $x\overline{I}_i\pi(x)$ for all $x \in X$. This is well defined because $\pi(X) \cap X = \emptyset$. Observe that $p^*(\overline{R}_i) = p^*(R_i)$ and $\overline{R}_i$ coincides with R_i^π on $A(\pi(X))$. Moreover, for all $\overline{x} \in \overline{X}$ and $i \in N$, there exists some $x_i \in X$ such that $\overline{x}\overline{R}_ix_i$: just take $x_i = \overline{x}$ if $\overline{x} \in X$ and $x_i = \pi^{-1}(\overline{x})$ if $\overline{x} \in \pi(X)$. Let $\overline{R}_N = (\overline{R}_1, ..., \overline{R}_n)$. Applying

[13] Contrast this with Gilboa, Samet and Schmeidler (2004) and Chambers and Hayashi (2006), where the appropriate normalization of the individual utilities is left unspecified and may vary arbitrarily across preference profiles. Note also that, because of the product form, the particular choice of 0-normalizations is inconsequential.

[14] This lemma is reminiscent of (but logically unrelated to) a result of Sen (1977): for social welfare *functionals* (i.e., mappings from profiles of *utility functions* into social orderings), the suitably defined conditions of Unrestricted Domain, Pareto Indifference, and Independence of Irrelevant Alternatives imply "Strong Neutrality".

Independence of Harmless Expansions to (4.1),

$$a\mathbf{R}(\overline{X}, \overline{R}_N)b. \tag{4.2}$$

Since $a^\pi \overline{I}_i a$ and $b^\pi \overline{I}_i b$ for all $i \in N$, Pareto Indifference implies $a^\pi \mathbf{I}(\overline{X}, \overline{R}_N)a$ and $b^\pi \mathbf{I}(\overline{X}, \overline{R}_N)b$. Hence from (4.2),

$$a^\pi \mathbf{R}(\overline{X}, \overline{R}_N)b^\pi. \tag{4.3}$$

For all $\overline{x} \in \overline{X}$ and $i \in N$, there exists some $x_i \in \pi(X)$ such that $\overline{x}\overline{R}_i x_i$. Since $\overline{R}_N$ coincides with R_N^π on $A(\pi(X))$, (4.3) and Independence of Harmless Expansions therefore imply $a^\pi \mathbf{R}(\pi(X), R_N^\pi)b^\pi$.

Step 2. $a^\pi \mathbf{R}(\pi(X), R_N^\pi)b^\pi$.

Choose $\rho \in \Pi(\mathbb{X})$ such that $\rho(X) \cap X = \rho(X) \cap \pi(X) = \emptyset$. This is possible because $\mathbb{X}$ is uncountable. By Step 1, (4.1) implies

$$a^\rho \mathbf{R}(\rho(X), R_N^\rho)b^\rho. \tag{4.4}$$

Next consider the bijection $\pi \circ \rho^{-1} \in \Pi(\mathbb{X})$. Since $(\pi \circ \rho^{-1})(\rho(X)) \cap \rho(X) = \emptyset$, Step 1 and (4.4) imply

$$(a^\rho)^{\pi\circ\rho^{-1}} \mathbf{R}((\pi \circ \rho^{-1})(\rho(X)), (R_N^\rho)^{\pi\circ\rho^{-1}})(b^\rho)^{\pi\circ\rho^{-1}}. \tag{4.5}$$

By definition, $(\pi \circ \rho^{-1})(\rho(X)) = \pi(X)$. Moreover, $(a^\rho)^{\pi\circ\rho^{-1}} = a^\pi$ since $(a^\rho)^{\pi\circ\rho^{-1}}(s) = (\pi \circ \rho^{-1})(a^\rho(s)) = (\pi \circ \rho^{-1})(\rho(a(s))) = \pi(a(s))$ for all $s \in S$. Likewise, $(b^\rho)^{\pi\circ\rho^{-1}} = b^\pi$ and $(R_N^\rho)^{\pi\circ\rho^{-1}} = R_N^\pi$. Hence (4.5) reduces to $a^\pi \mathbf{R}(\pi(X), R_N^\pi)b^\pi$. ■

Proof of the theorem. The proof of the "if" statement is straightforward. To prove the converse statement, fix a SWF $\mathbf{R}$ satisfying the Weak Pareto Principle, Independence of Harmless Expansions, and Continuity. This SWF satisfies Pareto Indifference, hence also Outcome Neutrality, by the above lemma.

For any $p_N \in \mathcal{P}^N$, define $\mathcal{D}(p_N) = \{(X, R_N) \in \mathcal{D} \mid p^*(R_N) = p_N\}$. This is the domain of problems in which the belief profile is p_N. Define the binary relations $\succ_{p_N}$, $\sim_{p_N}$, and $\succsim_{p_N}$ on $\mathbb{R}_+^N$ as follows: for all $v, w \in \mathbb{R}_+^N$,

(i) $v \succ_{p_N} w$ if and only if there exist $(X, R_N) \in \mathcal{D}(p_N)$, $U_N \in \mathcal{U}_0(X, R_N)$, and $a, b \in A(X)$ such that $U_N(a) = v$, $U_N(b) = w$, and $a\mathbf{P}(X, R_N)b$,

(ii) $v \sim_{p_N} w$ if and only if there exist $(X, R_N) \in \mathcal{D}(p_N)$, $U_N \in \mathcal{U}_0(X, R_N)$, and $a, b \in A(X)$ such that $U_N(a) = v$, $U_N(b) = w$, and $a\mathbf{I}(X, R_N)b$,

(iii) $v \succsim_{p_N} w$ if and only if $v \succ_{p_N} w$ or $v \sim_{p_N} w$.

Replacing $A(X)$ with X in statements (i) and (ii) yields an equivalent definition of the relations $\succ_{p_N}, \sim_{p_N}$, and $\succsim_{p_N}$, and we will freely use both definitions in the remainder of the proof. To check that the two formulations are indeed equivalent, fix $p_N \in \mathcal{P}^N$, $v, w \in \mathbb{R}_+^N$, and suppose there exist $(X, R_N) \in \mathcal{D}(p_N)$, $U_N \in \mathcal{U}_0(X, R_N)$, and $a, b \in A(X)$ such that $U_N(a) = v$, $U_N(b) = w$, and $a\mathbf{P}(X, R_N)b$ (respectively, $a\mathbf{I}(X, R_N)b$). We must

find $(X', R'_N) \in \mathcal{D}(p_N)$, $U'_N \in \mathcal{U}_0(X', R'_N)$, and $x, y \in X'$ such that $U'_N(x) = v$, $U'_N(y) = w$, and $x\mathbf{P}(X', R'_N)y$ (respectively, $x\mathbf{I}(X', R'_N)y$). Note that $U'_N(x)$, $U'_N(y)$ are well defined because of our convention to identify an outcome with the constant act assigning that outcome to every state of nature.

To do this, simply choose any two distinct outcomes $x, y \in \mathbb{X} \setminus X$ and define $X' = X \cup \{x, y\}$. For each $i \in N$, let R'_i be the Savage preference on $A(X')$ which coincides with R_i on $A(X)$ and is such that $xI'_i a$ and $yI'_i b$. Let U'_i be the Savage representation of R'_i which coincides with U_i on $A(X)$. Let $R'_N = (R'_1, ..., R'_n)$ and $U'_N = (U'_1, ..., U'_n)$. Because R'_N coincides with R_N on $A(X)$, we have $p^*(R'_N) = p_N$, hence $(X', R'_N) \in \mathcal{D}(p_N)$. Because $U_N \in \mathcal{U}_0(X, R_N)$ and $xI'_i a$ and $yI'_i b$ for all $i \in N$, we have $U'_N \in \mathcal{U}_0(X', R'_N)$ and $U'_N(x) = v$ and $U'_N(y) = w$. Finally, since for all $x' \in X'$ and $i \in N$ there is some $x_i \in X$ such that $x' R'_i x_i$, Independence of Harmless Expansions and Pareto Indifference imply $x\mathbf{I}(X', R'_N)a\mathbf{P}(X', R'_N)b\mathbf{I}(X', R'_N)y$ (respectively, $x\mathbf{I}(X', R'_N)a\mathbf{I}(X', R'_N)b\mathbf{I}(X', R'_N)y$), and we are done.

Step 1. *For all $p_N \in \mathcal{P}^N$, (i) the binary relation $\succsim_{p_N}$ is an ordering, and (ii) for all $v, w \in \mathbb{R}^N_+$ one and only one of the following statements holds: (a) $v \succ_{p_N} w$, (b) $w \succ_{p_N} v$, (c) $v \sim_{p_N} w$.*

Fix a belief profile $p_N \in \mathcal{P}^N$.

To prove reflexivity and completeness of $\succsim_{p_N}$, fix two (possibly equal) vectors $v, w \in \mathbb{R}^N_+$. Let $x_0, x_1, x_2, x_3 \in \mathbb{X}$ be four distinct outcomes and let $X = \{x_0, x_1, x_2, x_3\}$. For each $i \in N$, choose a number $z_i \in \mathbb{R}_+$ such that $z_i \neq v_i, w_i$, define $u_i : X \to \mathbb{R}_+$ by $u_i(x_0) = 0$, $u_i(x_1) = v_i$, $u_i(x_2) = w_i$, and $u_i(x_3) = z_i$. Define $U_i : A(X) \to \mathbb{R}_+$ by $U_i(a) = U(a, u_i, p_i) = \int_S (u_i \circ a) dp_i$ for all $a \in A(X)$, and let R_i be the preference on $A(X)$ represented by U_i : by construction, $R_i \in \mathcal{R}^*(X)$ and $p^*(R_i) = p_i$. Letting $U_N := (U_1, ..., U_n)$ and $R_N = (R_1, ..., R_n)$, we have $(X, R_N) \in \mathcal{D}(p_N)$ and $U_N \in \mathcal{U}_0(X, R_N)$. Since $\mathbf{R}(X, R_N)$ is complete and reflexive, we must have $x_1\mathbf{R}(X, R_N)x_2$ or $x_2\mathbf{R}(X, R_N)x_1$. Since $U_N(x_1) = v$ and $U_N(x_2) = w$, we have $v \succsim_{p_N} w$ or $w \succsim_{p_N} v$.

To prove transitivity of $\succsim_{p_N}$, fix $v^1, v^2, v^3 \in \mathbb{R}^N_+$ such that $v^1 \succsim_{p_N} v^2 \succsim_{p_N} v^3$. By definition, there exist $(X^1, R^1_N), (X^2, R^2_N) \in \mathcal{D}(p_N)$, $U^1_N \in \mathcal{U}_0(X^1, R^1_N)$, $U^2_N \in \mathcal{U}_0(X^2, R^2_N)$, $x^1, y^1 \in X^1$, and $x^2, y^2 \in X^2$ such that

$$U^1_N(x^1) = v^1,\ U^1_N(y^1) = v^2 = U^2_N(x^2),\ \text{and}\ U^2_N(y^2) = v^3, \tag{4.6}$$

and

$$x^1\mathbf{R}(X^1, R^1_N)y^1 \text{ and } x^2\mathbf{R}(X^2, R^2_N)y^2. \tag{4.7}$$

By Outcome Neutrality and because $\mathbb{X}$ is uncountable, we may assume that $X^1 \cap X^2 = \emptyset$. Let $X = X^1 \cup X^2$. For each $i \in N$, define $u_i : X \to \mathbb{R}_+$ by

$$u_i(x) = \begin{cases} U^1_i(x) \text{ if } x \in X^1, \\ U^2_i(x) \text{ if } x \in X^2, \end{cases} \tag{4.8}$$

and define $U_i : A(X) \to \mathbb{R}_+$ by $U_i(a) = U(a, u_i, p_i) = \int_S (u_i \circ a) dp_i$ for all $a \in A(X)$. Let R_i be the Savage preference on $A(X)$ represented by U_i, let $U_N = (U_1, ..., U_n)$, and let $R_N = (R_1, ..., R_n)$.

Note that R_N coincides with R_N^1 on $A(X^1)$ and with R_N^2 on $A(X^2)$. Moreover, because $U_N^1 \in \mathcal{U}_0(X^1, R_N^1)$ and $U_N^2 \in \mathcal{U}_0(X^2, R_N^2)$, (4.8) implies that $U_N \in \mathcal{U}_0(X, R_N)$. Moreover, for all $x \in X$ and $i \in N$, there exist $x_i^1 \in X^1$, $x_i^2 \in X^2$ such that $xR_ix_i^1$, $xR_ix_i^2$. We may therefore apply Independence of Harmless Expansions to (4.7) and conclude

$$x^1 \mathbf{R}(X, R_N) y^1 \text{ and } x^2 \mathbf{R}(X, R_N) y^2.$$

On the other hand, (4.6) and (4.8) imply $y^1 I_i x^2$ for all $i \in N$, hence by Pareto Indifference,

$$y^1 \mathbf{I}(X, R_N) x^2.$$

Transitivity of $\mathbf{R}(X, R_N)$ now implies $x^1 \mathbf{R}(X, R_N) y^2$. Since $(X, R_N) \in \mathcal{D}(p_N)$, $U_N \in \mathcal{U}_0(X, R_N)$, and $U_N(x^1) = v^1$ and $U_N(y^2) = v^3$, the definition of $\succsim_{p_N}$ gives us $v^1 \succsim_{p_N} v^3$. This establishes the transitivity of $\succsim_{p_N}$ and completes the proof of statement (i) in Step 1.

The proof of statement (ii) is similar to the proof of transitivity and omitted for brevity.

Step 2. *For all $p_N \in \mathcal{P}^N$ there exists $\gamma \in \Delta_N$ such that, for all $v, w \in \mathbb{R}_+^N$, $v \succsim_{p_N} w \Leftrightarrow \prod_{i \in N} v_i^{\gamma_i} \geq \prod_{i \in N} w_i^{\gamma_i}$.*

Note that the number γ in the above statement may vary with p_N. To prove that statement, fix $p_N \in \mathcal{P}^N$. In order to alleviate notation, we write $\succ$, $\sim$, and $\succsim$ instead of $\succ_{p_N}$, $\sim_{p_N}$, and $\succsim_{p_N}$. We use $\geq, >, \gg$ to write inequalities in $\mathbb{R}_+^N$.

We begin by establishing three properties of $\succsim$. The first, scale invariance, is key to the multiplicative form of the SWFs satisfying our axioms.

Step 2.1. $\succsim$ is *scale invariant*: $v \succsim w \Leftrightarrow \lambda * v \succsim \lambda * w$ *for all* $\lambda \in \mathbb{R}_{++}^N$, *where* $\lambda * v = (\lambda_1 v_1, ..., \lambda_n v_n)$.

To check this point, fix $v, w \in \mathbb{R}_+^N$, $\lambda \in \mathbb{R}_{++}^N$, and suppose $v \succ w$ (respectively, $v \sim w$). By definition, there exist $(X, R_N) \in \mathcal{D}(p_N)$, $U_N \in \mathcal{U}_0(X, R_N)$, and $a, b \in A(X)$ such that $U_N(a) = v$, $U_N(b) = w$, and $a\mathbf{P}(X, R_N)b$ (respectively, $a\mathbf{I}(X, R_N)b$). For each $i \in N$, define $V_i : A(X) \to \mathbb{R}_+$ by $V_i(c) = \lambda_i U_i(c)$ for all $c \in A(X)$, and let $V_N = (V_1, ..., V_n)$. Observe that $V_N \in \mathcal{U}_0(X, R_N)$ and $V_N(a) = \lambda * v$, $V_N(b) = \lambda * w$. Since $(X, R_N) \in \mathcal{D}(p_N)$ and $a\mathbf{P}(X, R_N)b$ (respectively, $a\mathbf{I}(X, R_N)b$), the definition of $\succ$ (respectively, $\sim$) implies $\lambda * v \succ \lambda * w$ (respectively, $\lambda * v \sim \lambda * w$), as desired.

Step 2.2. $\succsim$ is *weakly monotonic*: $v \gg w \Rightarrow v \succ w$.

This follows immediately from the fact that $\mathbf{R}$ satisfies (part (i) of) the Weak Pareto Principle.

Step 2.3. $\succsim$ *is continuous: for all $u, v \in \mathbb{R}_+^N$ and every sequence (u^t) in $\mathbb{R}_+^N$ converging to u, (i) if $u^t \succsim v$ for all $t \in \mathbb{N}$, then $u \succsim v$, and (ii) if $v \succsim u^t$ for all $t \in \mathbb{N}$, then $v \succsim u$.*

This follows from Continuity and Independence of Harmless Expansions. Fix $u, v \in \mathbb{R}^N_+$ and a sequence (u^t) in $\mathbb{R}^N_+$ converging to u. We only prove statement (i); the proof of (ii) is the same, mutatis mutandis.

Suppose $u^t \succsim v$ for all $t \in \mathbb{N}$. For each $t \in \mathbb{N}$, there exist $(X^t, R^t_N) \in \mathcal{D}(p_N)$, $U^t_N \in \mathcal{U}_0(X^t, R^t_N)$, and $x^t, y^t \in X^t$ such that $U^t_N(x^t) = u^t$, $U^t_N(y^t) = v$, and

$$x^t \mathbf{R}(X^t, R^t_N) y^t. \tag{4.9}$$

Moreover, there exist $(X^0, R^0_N) \in \mathcal{D}(p_N)$, $U^0_N \in \mathcal{U}_0(X^0, R^0_N)$, and $x^0, y^0 \in X^0$ such that $U^0_N(x^0) = u$ and $U^0_N(y^0) = v$.

By Outcome Neutrality and because $\mathbb{X}$ is uncountable, we may assume that $X^t \cap X^{t'} = \emptyset$ for all distinct $t, t' \in \mathbb{N} \cup \{0\}$. Define $\overline{X} = \cup_{t \in \mathbb{N} \cup \{0\}} X^t$. For each $i \in N$, define $\overline{u}_i : \overline{X} \to \mathbb{R}$ by

$$\overline{u}_i(x) = U_i^{t(x)}(x), \tag{4.10}$$

where $t(x)$ is the unique integer $t \in \mathbb{N} \cup \{0\}$ such that $x \in X^t$. Define $\overline{U}_i : A(\overline{X}) \to \mathbb{R}$ by $\overline{U}_i(a) = \int_S (\overline{u}_i \circ a) dp_i$ for all $a \in A(\overline{X})$. Let $\overline{R}_i$ be the Savage preference on $A(\overline{X})$ represented by $\overline{U}_i$, let $\overline{U}_N = (\overline{U}_1, ..., \overline{U}_n)$, and let $\overline{R}_N = (\overline{R}_1, ..., \overline{R}_n)$. Note that $\overline{R}_N$ coincides with R^t_N on $A(X^t)$ for each $t \in \mathbb{N} \cup \{0\}$. Moreover, because $U^t_N \in \mathcal{U}_0(X^t, R^t_N)$ for each $t \in \mathbb{N} \cup \{0\}$, (4.10) implies that $\overline{U}_N \in \mathcal{U}_0(\overline{X}, \overline{R}_N)$. Moreover, for all $\overline{x} \in \overline{X}$, all $t \in \mathbb{N} \cup \{0\}$, and all $i \in N$, there is some $x^t_i \in X^t$ such that $\overline{x} \overline{R}_i x^t_i$. Applying Independence of Harmless Expansions to (4.9), we get

$$x^t \mathbf{R}(\overline{X}, \overline{R}_N) y^t \text{ for all } t \in \mathbb{N}. \tag{4.11}$$

Since $\overline{U}_N(y^t) = v$ for all $t \in \mathbb{N} \cup \{0\}$, we have $y^t \overline{I}_i y^0$ for all $t \in \mathbb{N} \cup \{0\}$ and all $i \in N$. From (4.11) and Pareto Indifference,

$$x^t \mathbf{R}(\overline{X}, \overline{R}_N) y^0 \text{ for all } t \in \mathbb{N}. \tag{4.12}$$

Since $\overline{U}_N(x^t) = u^t \to u = \overline{U}_N(x^0)$, we have $U_N(x^t) \to U_N(x^0)$ for all $U_N \in \mathcal{U}(\overline{X}, \overline{R}_N)$. That is, the sequence (x^t) in $A(\overline{X})$ converges to x^0 with respect to $\overline{R}_N$. Because $\mathbf{R}$ satisfies Continuity, (4.12) now implies $x^0 \mathbf{R}(\overline{X}, \overline{R}_N) y^0$. Since $(\overline{X}, \overline{R}_N) \in \mathcal{D}(p_N)$, $\overline{U}_N \in \mathcal{U}_0(\overline{X}, \overline{R}_N)$, and $\overline{U}_N(x^0) = u$ and $\overline{U}_N(y^0) = v$, the definition of $\succsim$ yields $u \succsim v$.

Step 2.4. *There exists $\gamma \in \Delta_N$ such that, for all $v, w \in \mathbb{R}^N_+$, $v \succsim w \Leftrightarrow \prod_{i \in N} v_i^{\gamma_i} \geq \prod_{i \in N} w_i^{\gamma_i}$.*

Since $\succsim$ is continuous, it admits a continuous numerical representation: there exists a continuous function $W : \mathbb{R}^N_+ \to \mathbb{R}$ such that, for all $v, w \in \mathbb{R}^N_+$, $v \succsim w \Leftrightarrow W(v) \geq W(w)$. Because $\succsim$ is scale-invariant and weakly monotonic, we have that for all $v, w \in \mathbb{R}^N_+$ and all $\lambda \in \mathbb{R}^N_{++}$, (i) $W(v) \geq W(w) \Leftrightarrow W(\lambda * v) \geq W(\lambda * w)$ and (ii) $v \gg w \Rightarrow W(v) > W(w)$ (hence also (iii) $v \geq w \Rightarrow W(v) \geq W(w)$ because W is continuous).

By a theorem of Osborne (1976), properties (i) and (iii) imply that there exist non-negative real numbers $\gamma_1, ..., \gamma_n$ and an increasing function $g : \mathbb{R} \to \mathbb{R}$ such that $W(v) = g(\prod_{i \in N} v_i^{\gamma_i})$ for all $v \in \mathbb{R}^N_+$. Since in our case W also satisfies (ii), not all the numbers $\gamma_1, ..., \gamma_n$

can be zero, and we may assume without loss of generality that $\gamma = (\gamma_1, ..., \gamma_n) \in \Delta_N$. Since W represents $\succsim$, we have $v \succsim w \Leftrightarrow \prod_{i \in N} v_i^{\gamma_i} \geq \prod_{i \in N} w_i^{\gamma_i}$ for all $v, w \in \mathbb{R}_+^N$. This completes the proof of Step 2.

Since p_N was arbitrary in the argument above, we have proved that there exists a function $\gamma : \mathcal{P}^N \to \Delta_N$ such that, for every $p_N \in \mathcal{P}^N$ and all $v, w \in \mathbb{R}_+^N$,

$$v \succsim_{p_N} w \Leftrightarrow \prod_{i \in N} v_i^{\gamma_i(p_N)} \geq \prod_{i \in N} w_i^{\gamma_i(p_N)}.$$

Step 3. *For all $(X, R_N) \in \mathcal{D}$ and all $a, b \in A(X)$, we have*

$$a\mathbf{R}(X, R_N)b \Leftrightarrow \left[\prod_{i \in N} U_i(a)^{\gamma_i(p^*(R_N))} \geq \prod_{i \in N} U_i(b)^{\gamma_i(p^*(R_N))} \textit{ for all } U_N \in \mathcal{U}_0(X, R_N)\right]$$

Fix $(X, R_N) \in \mathcal{D}$ and $a, b \in A(X)$. If $\prod_{i \in N} U_i(a)^{\gamma_i(p^*(R_N))} \geq \prod_{i \in N} U_i(b)^{\gamma_i(p^*(R_N))}$ for all $U_N \in \mathcal{U}_0(X, R_N)$, Step 2 implies $U_N(a) \succsim_{p^*(R_N)} U_N(b)$ for all $U_N \in \mathcal{U}_0(X, R_N)$, and the definition of $\succsim_{p^*(R_N)}$ (and the fact that it is an ordering) implies that $a\mathbf{R}(X, R_N)b$. Conversely, if $\prod_{i \in N} U_i(b)^{\gamma_i(p^*(R_N))} > \prod_{i \in N} U_i(a)^{\gamma_i(p^*(R_N))}$ for some $U_N \in \mathcal{U}_0(X, R_N)$, Step 2 implies $U_N(b) \succ_{p^*(R_N)} U_N(a)$ and the definition of $\succsim_{p^*(R_N)}$ implies $b\mathbf{P}(X, R_N)a$.■

A general comment is in order about the proof. Because the argument given above works for any fixed profile of beliefs, the theorem remains true on the subdomain of preference profiles over acts where the beliefs of all agents are the same. This means that the result can be reformulated in the context of the aggregation of von Neumann-Morgenstern preferences over lotteries: properly rewritten, our axioms characterize the (exogenously) weighted 0-normalized Nash SWFs in that context, and Anonymity easily pins down the uniform 0-normalized Nash SWF. It is instructive to compare this variant of our result with Dhillon and Mertens' (1999) characterization of relative utilitarianism: dropping the requirement that society's preference be von Neumann-Morgenstern and strengthening Independence of Redundant Alternatives to Independence of Harmless Alternatives leads us to give up relative utilitarianism in favor of 0-normalized Nash welfarism.

5. Discussion

This section addresses three criticisms that may be formulated against the belief-weighted Nash SWFs.

5.1. Lack of rationality?

In this subsection, we fix the set of relevant outcomes X and drop it from our notation whenever there is no risk of confusion. Thus, we write A instead of $A(X)$ for the set of relevant acts, and a problem (X, R_N) reduces to its profile component R_N.

Perhaps the main drawback of a belief-weighted Nash SWF (henceforth simply called a Nash SWF) is that the social preference it recommends (henceforth called a Nash preference) may not be of the Savage type. More precisely: if $\mathbf{R}$ is a Nash SWF that never puts weight one on a single agent's belief (i.e., $\gamma(p^*(R_N)) \ll 1$ for all profiles R_N), then there exist profiles R_N where $\mathbf{R}(R_N)$ is not a Savage preference. This is problematic because Savage's axioms are generally regarded as criteria of rationality in the face of uncertainty.

In order to assess the severity of this problem, it is important to examine (i) which of Savage's axioms may be violated by a Nash preference, and (ii) whether these axioms are compelling for a *social* preference.

Let us begin by recalling Savage's axiomatic system. Let R be a preference relation on A. For all $a, b \in A$ and $E \subseteq S$, define the act $a_E b$ by $(a_E b)(s) = a(s)$ if $s \in E$ and $(a_E b)(s) = b(s)$ otherwise. Call an event E R-null if $a_E c \; I \; b_E c$ for all $a, b, c \in A$, and let $\mathcal{E}^+(R)$ be the set of events which are not R-null. For all $E \subseteq S$, define the relation R_E on A by $aR_E b \Leftrightarrow [aRb$ and $a(s) = b(s)$ for all $s \in S \setminus E]$. Savage's axioms are:

(P1) R is an ordering,

(P2) for all $a, b, c, c' \in A$ and all $E \subseteq S$, $a_E c \; R \; b_E c \Leftrightarrow a_E c' \; R \; b_E c'$,

(P3) for all $x, y \in X$, all $a \in A$, and all $E \in \mathcal{E}^+(R)$, $xRy \Leftrightarrow x_E a \; R \; y_E a$,

(P4) for all $x, x', y, y' \in X$ such that xPy and $x'Py'$, and for all $E, E' \subseteq S$, $x_E y \; R \; x_{E'} y \Leftrightarrow x'_E y' \; R \; x'_{E'} y'$,

(P5) there exist $a, b \in A$ such that aPb,

(P6) for all $a, b, c \in A$ such that aPb, there is a partition $\{E_1, ..., E_n\}$ of S such that $c_{E_i} a \; P \; b$ and $a \; P \; c_{E_i} b$ for $i = 1, ..., n$,

(P7) for all $a, b \in A$ and all $E \subseteq S$, (i) $[aP_E b(s)$ for all $s \in E] \Rightarrow aR_E b$ and (ii) $[b(s)P_E a$ for all $s \in E] \Rightarrow bR_E a$.

Let us examine which of these axioms may be violated by a Nash preference. For simplicity, let us assume that $N = \{1, 2\}$ and let $\mathbf{R}$ be the *uniform* Nash SWF ranking acts according to the (uniformly weighted) product of their $(0, 1)$-normalized utilities.

The very basic axiom (P1) poses no problem: $\mathbf{R}(R_{\{1,2\}})$ is an ordering for every profile $R_{\{1,2\}}$. One can also check that the somewhat technical conditions (P5) to (P7) are satisfied by $\mathbf{R}(R_{\{1,2\}})$ at every $R_{\{1,2\}}$.[15]

On the other hand, $\mathbf{R}(R_{\{1,2\}})$ violates each of the separability axioms (P2) to (P4) at some profile $R_{\{1,2\}}$. To see that (P2) need not hold, let $x, y \in X$, $E \subseteq S$, and suppose $R_{\{1,2\}}$ is such that $p^*(R_i)(E) = \frac{1}{2}$ for $i = 1, 2$ and

$$\begin{aligned} U^*(x, R_1) &= 1, \; U^*(y, R_1) = 0, \\ U^*(x, R_2) &= 0, \; U^*(y, R_2) = 1. \end{aligned}$$

[15] The preference $\mathbf{R}(R_{\{1,2\}})$ satisfies (P5) because the function $(w_1, w_2) \mapsto u_1 u_2$ is strictly quasi-concave on $(0, 1]^2$.

Then $\prod_{i=1,2} U^*(x, R_i) = \prod_{i=1,2} U^*(y, R_i) = 0$ and $\prod_{i=1,2} U^*(x_E y, R_i) = \prod_{i=1,2} U^*(y_E x, R_i) = \frac{1}{4}$, so that $x_E y \ \mathbf{P}(R_{\{1,2\}})\ y_E y = y$ and $y_E x \ \mathbf{P}(R_{\{1,2\}})\ x_E x = x$, a violation of (P2) (with $c = y$ and $c' = x$).

Since $\mathbf{R}$ satisfies the Weak Pareto Principle, we know from Chambers and Hayashi (2006) that violations of (P3) and (P4) must occur. For the sake of completeness, we quickly provide examples of such violations.

To see that (P3) may fail, let $x, y, z \in X$, $E \subseteq S$, and suppose $R_{\{1,2\}}$ is such that $p^*(R_i)(E) = \frac{1}{2}$ for $i = 1, 2$ and

$$\begin{aligned} U^*(y, R_1) &= 0,\ U^*(x, R_1) = \frac{1}{2},\ U^*(z, R_1) = 1, \\ U^*(y, R_2) &= 1,\ U^*(x, R_2) = \frac{1}{2},\ U^*(z, R_2) = 0. \end{aligned}$$

Then $\prod_{i=1,2} U^*(x, R_i) = \frac{1}{4} > \prod_{i=1,2} U^*(y, R_i) = 0$ and $\prod_{i=1,2} U^*(x_E z, R_i) = \frac{3}{16} < \prod_{i=1,2} U^*(y_E z,\ R_i) = \frac{1}{4}$, so that $x \mathbf{P}(R_{\{1,2\}}) y$ and $y_E z \ \mathbf{P}(R_{\{1,2\}})\ x_E z$. Since the latter preference implies that $E \in \mathcal{E}^+(\mathbf{R}(R_{\{1,2\}}))$, $\mathbf{R}(R_{\{1,2\}})$ violates (P3).

For a violation of (P4), let $x, y, y' \in X$, $E \subseteq S$, and suppose $R_{\{1,2\}}$ is such that $p^*(R_1)(E) = \frac{1}{3}$, $p^*(R_2)(E) = \frac{2}{3}$, and

$$\begin{aligned} U^*(y, R_1) &= 1,\ U^*(x, R_1) = \frac{1}{2},\ U^*(y', R_1) = 0, \\ U^*(y, R_2) &= 0,\ U^*(x, R_2) = \frac{1}{2},\ U^*(y', R_2) = 1. \end{aligned}$$

Letting $E' = S \setminus E$, we have $\prod_{i=1,2} U^*(x_E y, R_i) = (\frac{5}{6})(\frac{1}{3}) > \prod_{i=1,2} U^*(x_{E'} y, R_i) = (\frac{2}{3})(\frac{1}{6})$ and $\prod_{i=1,2} U^*(x_E y', R_i) = (\frac{1}{6})(\frac{2}{3}) < \prod_{i=1,2} U^*(x_{E'} y', R_i) = (\frac{1}{3})(\frac{5}{6})$, so that $x_E y \ \mathbf{P}(R_{\{1,2\}})\ x_{E'} y$ and $x_{E'} y' \ \mathbf{P}(R_{\{1,2\}})\ x_E y'$. But $\prod_{i=1,2} U^*(y, R_i) = \prod_{i=1,2} U^*(y', R_i) = 0 < \prod_{i=1,2} U^*(x, R_i)$, a violation of (P4) (with $x' = x$).

Savage proposed axioms (P1) to (P7) as rationality criteria applicable to an *individual* preference relation. We would like to argue that the relevance of these axioms should be reexamined if R is a *social* preference. The reason is that such a preference is not a primitive concept; rather, it is constructed from a profile of individual preferences through a SWF: $R = \mathbf{R}(R_N)$. Therefore,

(a) the appeal of an axiom imposed on a social preference $R = \mathbf{R}(R_N)$ may well depend upon the preference profile R_N that R summarizes, and

(b) the preference profile R_N itself may have to enter into the proper formulation of an axiom imposed on the social preference $\mathbf{R}(R_N)$.

As an elementary illustration of claim (a), suppose $X = \{x, y\}$ and $N = \{1, 2\}$. If $R_{\{1,2\}}$ is a profile such that xP_1y, yP_2x, and $p^*(R_1) = p^*(R_2)$, it is unclear whether (P5) should be imposed on $\mathbf{R}(R_{\{1,2\}})$. Note in particular that the popular relative utilitarian SWF (which ranks acts according to the sum of the $(0, 1)$-normalized utilities they generate) indeed deems all acts equally good at $R_{\{1,2\}}$.

By contrast, if $R'_{\{1,2\}}$ is a unanimous profile (that is, $R'_1 = R'_2$), then (P5) is totally compelling for $\mathbf{R}(R'_{\{1,2\}})$ because the preference aggregation is trivial and society should behave as a single individual. Note that any Paretian SWF $\mathbf{R}$ indeed recommends $\mathbf{R}(R'_{\{1,2\}}) = R'_1 = R'_2$, which of course is a Savage preference.

We focused on (P5) for simplicity but, as we will see later, claim (a) also applies to the other Savage axioms.

We now come to claim (b). Savage's separability axioms (P2) to (P4) are restrictions of the following type: "if an act a is weakly preferred to an act b, and if acts a', b' are suitably related to a, b, then a' should be weakly preferred to b'". We claim that in a social decision context, simply knowing that "a is weakly preferred to b" may not be informative enough to conclude that "a' should also be weakly preferred to b'". Indeed, a social preference summarizes an entire *profile* of individual preferences; this profile is the *reason* why a is weakly preferred to b and it may not be wise to ignore it when deciding whether society should also weakly prefer a' to b'.

Consider our earlier example of a violation of (P2). Given the profile $R_{\{1,2\}}$ described in that example, a sensible motivation for the social preference $a_E b \ \mathbf{P}(R_{\{1,2\}})\ b$ is that act $a_E b$ offers both individuals a chance to get their favorite outcome whereas b yields individual $1'$s favorite outcome –which is also 2's worst– in all states. Ex-ante, $a_E b$ appears more equitable than b.[16] But the social preference $a\ \mathbf{P}(R_{\{1,2\}})\ b_E a$, which (P2) then prescribes, cannot be justified on similar grounds –indeed, the opposite preference is supported by the same fairness considerations. This suggests that (P2) may not be a reasonable restriction on $\mathbf{R}(R_{\{1,2\}})$.

On the other hand, at a profile $R'_{\{1,2\}}$ where $p^*(R'_i)(E) = \frac{1}{2}$ for $i = 1, 2$ and $U^*(a, R'_i) = 1$, $U^*(b, R'_i) = 0$ for $i = 1, 2$, the requirement $a_E b\ \mathbf{P}(R'_{\{1,2\}})\ b_E b \Leftrightarrow a_E a\ \mathbf{P}(R'_{\{1,2\}})\ b_E a$ is completely natural –and indeed satisfied by any Nash SWF $\mathbf{R}$. A plausible reason for the social preference $a_E b\ \mathbf{P}(R'_{\{1,2\}})\ b_E b$ is that both individuals prefer the former act to the latter, and this same reason justifies the preference $a_E a\ \mathbf{P}(R'_{\{1,2\}})\ b_E a$.

This discussion again backs claim (a) –but also claim (b): the preference profile itself may have to enter the premise of an axiom bearing on the social preference that summarizes it. In that spirit, here is a variant of (P2) which the Nash SWFs satisfy.

(WP2) for all $a, b, c, c' \in A$, all $E \subseteq S$, and all R_N such that $cI_i c'$ for all $i \in N$, $a_E c\ \mathbf{R}(R_N)\ b_E c \Leftrightarrow a_E c'\ \mathbf{R}(R_N)\ b_E c'$.

This is admittedly a much weaker separability condition but we believe it lies on safer grounds than (P2). The restriction assumed on R_N is meant to guarantee that the reason behind the social preference between $a_E c$ and $b_E c$ also justifies the preference between $a_E c'$ and $b_E c'$.

Axiom (P4) admits a particularly interesting variant. Denote by $W(R_i) = \{x \in X \mid yR_i x \text{ for all } y \in X\}$ the set of worst outcomes according to the preference R_i, and consider

[16] This well-known argument is a variant of Diamond's (1967) criticism of utilitarianism –applied to social preferences over acts rather than lotteries.

the following requirement:

(WP4) for all $x, x', y, y' \in X$, all $E, E' \subseteq S$, and all R_N such that $x, x' \notin \cup_{i\in N} W(R_i)$ and $y, y\prime \in \cap_{i\in N} W(R_i)$, $x_E y \ \mathbf{R}(R_N)\ x_{E'} y \Leftrightarrow x'_E y' \ \mathbf{R}(R_N)\ x'_{E'} y'$.

An outcome in $\cap_{i\in N} W(R_i)$ is everybody's worst: call it *bad* (at R_N). An outcome not in $\cup_{i\in N} W(R_i)$ is nobody's worst: call it *good* (at R_N). Axiom (WP4) requires that society's preference between two binary acts that yield either a good outcome or a bad outcome should be independent of the specification of these two outcomes.

A SWF $\mathbf{R}$ satisfying (WP4) generates at each profile R_N a well-defined likelihood relation $\succsim (R_N, \mathbf{R})$ on the set of events:

$$E \ \succsim (R_N, \mathbf{R})\ E' \Leftrightarrow x_E y \ \mathbf{R}(R_N)\ x_{E'} y \text{ for all } x \notin \cup_{i\in N} W(R_i) \text{ and } y \in \cap_{i\in N} W(R_i). \quad (5.1)$$

The interpretation of the statement $E \succsim (R_N, \mathbf{R})\ E'$ is that, at R_N, society believes that event E is at least as likely as event E' : this interpretation makes sense because society *always* weakly prefers a good outcome if E occurs and a bad outcome otherwise to a good outcome if E' occurs and a bad outcome otherwise.

If $\mathbf{R}$ satisfies (WP4), the binary relation $\succsim (R_N, \mathbf{R})$ is transitive for *every* profile R_N.[17] Of course, at a profile R_N where no good outcome exists ($\cup_{i\in N} W(R_i) = X$) or no bad outcome exists ($\cap_{i\in N} W(R_i) = \varnothing$), (5.1) implies that $E \sim (R_N, \mathbf{R})\ E'$ for all E, E': society's belief is completely indeterminate.

Interestingly, the Nash SWFs satisfy WP4. To see this, consider for simplicity the uniform Nash SWF $\mathbf{R}$ and observe that if x is a good outcome at R_N and y is a bad outcome at R_N, then for any event E,

$$\prod_{i\in N} U^*(x_E y, R_i) = \prod_{i\in N} p^*(R_i)(E) U^*(x, R_i)$$

because $U^*(y, R_i) = 0$ for all $i \in N$. It follows that for any good outcomes x, x', any bad outcomes y, y', and any two events E, E',

$$x_E y \mathbf{R}(R_N) x_{E'} y \Leftrightarrow \prod_{i\in N} p^*(R_i)(E) \geq \prod_{i\in N} p^*(R_i)(E'), \quad (5.2)$$

and

$$x'_E y' \mathbf{R}(R_N) x'_{E'} y' \Leftrightarrow \prod_{i\in N} p^*(R_i)(E) \geq \prod_{i\in N} p^*(R_i)(E'),$$

hence,

$$x_E y \ \mathbf{R}(R_N)\ x_{E'} y \Leftrightarrow x'_E y' \mathbf{R}(R_N) x'_{E'} y',$$

as desired.

[17] This follows directly from the transitivity of the social preference relation $\mathbf{R}(R_N)$: for any events E, E', E'' such that $E \succsim (R_N, \mathbf{R})\ E' \succsim (R_N, \mathbf{R})\ E''$ we have $x_E y \ \mathbf{R}(R_N)\ x_{E'} y \ \mathbf{R}(R_N)\ x_{E''} y$, hence $x_E y \ \mathbf{R}(R_N)\ x_{E''} y$, for every good outcome x at R_N and every bad outcome y at R_N, implying $E \succsim (R_N, \mathbf{R})\ E''$.

At any profile R_N where a good outcome and a bad outcome exist, the social likelihood relation generated by the uniform Nash SWF $\mathbf{R}$ takes a simple form. From (5.1) and (5.2),

$$E \succsim (R_N, \mathbf{R})\ E' \Leftrightarrow \prod_{i \in N} p^*(R_i)(E) \geq \prod_{i \in N} p^*(R_i)(E'). \tag{5.3}$$

The social likelihood of an event is the product of the probabilities attached to it by the individuals. Society's belief $\succsim (R_N, \mathbf{R})$ cannot generally be represented by a probability measure but it is an ordering on 2^S.

A subtle aspect of the above construction is that two social beliefs $\succsim (R_N, \mathbf{R})$, $\succsim (R'_N, \mathbf{R})$ may differ *even if the social preferences* $\mathbf{R}(R_N)$, $\mathbf{R}(R'_N)$ *coincide*. A social belief is not a property of a social preference *per se*; it depends explicitly on the profile that generates this preference through the SWF.

As an illustration, suppose $N = \{1, 2\}$, $X = \{x, y\}$, and $\mathbf{R}$ is the uniform Nash SWF. Consider first a unanimous profile $R_{\{1,2\}}$. In such a profile, agents have identical beliefs: $p^*(R_1) = p^*(R_2) = p$. Suppose xR_iy for $i = 1, 2$. Since X only contains the two outcomes x, y, an act a is completely described by the event $E_a = \{s \in S \mid a(s) = x\}$. For any two acts a, b, we have $aR_ib \Leftrightarrow p(E_a) \geq p(E_b)$ for $i = 1, 2$ and the social preference at $R_{\{1,2\}}$ coincides with the common preference of the individuals:

$$a\mathbf{R}(R_{\{1,2\}})b \Leftrightarrow p(E_a) \geq p(E_b).$$

Using (5.3), the social likelihood relation generated by $\mathbf{R}$ at $R_{\{1,2\}}$ is given by

$$E \succsim (R_{\{1,2\}}, \mathbf{R})\ E' \Leftrightarrow p(E) \geq p(E'),$$

that is, society's belief coincides with the common probabilistic belief of its members.

Consider next a profile $R'_{\{1,2\}}$ where individuals have "opposite beliefs" *and* "opposite tastes": $p^*(R'_1)(E) = p^*(R'_2)(S \setminus E) = p(E)$ for every event E, $U^*(x, R'_1) = 1$, $U^*(y, R'_1) = 0$, and $U^*(x, R'_2) = 0$, $U^*(y, R'_2) = 1$. For all acts a, b, the social preference at $R_{\{1,2\}}$ is given by

$$\begin{aligned} a\mathbf{R}(R_{\{1,2\}})b &\Leftrightarrow U^*(a, R'_1)U^*(a, R'_2) \geq U^*(b, R'_1)U^*(b, R'_2) \\ &\Leftrightarrow [p^*(R'_1)(E_a)]\,[p^*(R'_2)(S \setminus E_a)] \geq [p^*(R'_1)(E_b)]\,[p^*(R'_2)(S \setminus E_b)] \\ &\Leftrightarrow [p(E_a)]^2 \geq [p(E_b)]^2 \\ &\Leftrightarrow p(E_a) \geq p(E_b), \end{aligned}$$

that is, $\mathbf{R}(R_{\{1,2\}}) = \mathbf{R}(R_{\{1,2\}})$. But since $W(R'_1) \cap W(R'_2) = \varnothing$ and $W(R'_1) \cup W(R'_2) = X$ the social likelihood relation generated by $\mathbf{R}$ at $R_{\{1,2\}}$ is indeterminate:

$$E \sim (R_{\{1,2\}}, \mathbf{R})\ E' \text{ for all } E, E' \subseteq S,$$

so that $\succsim (R_{\{1,2\}}, \mathbf{R}) \neq \succsim (R_{\{1,2\}}, \mathbf{R})$.

Upon reflection, this does seem right and (WP4) may be a good alternative to imposing (P4) on the social preference at all profiles. Imposing Savage's axioms at every profile

not only yields a well-defined probabilistic social belief, but also forces that belief to be the same at all profiles that generate the same social preference. However, as the above example shows, a given social preference may summarize two preference profiles with radically different belief components –provided these differences are "offset" by countervailing differences in tastes. Clearly, one cannot expect the same social belief to aggregate both of these belief profiles well.

To conclude, let us briefly take stock:
(i) the social preference recommended by a Nash SWF at a given profile always satisfies Savage's fundamental axiom (P1) but may violate each of (P2) to (P4) –it does satisfy all axioms when that profile is unanimous;
(ii) there are profiles where it may not be wise to expect the social preference to satisfy axioms (P2) to (P4);
(iii) profile-dependent variants of Savage's axioms can be defined –the Nash SWFs satisfy a variant of (P4) guaranteeing a well-defined, non-degenerate (but non-probabilistic) social belief at all profiles where a "good" outcome and a "bad" outcome exist.

5.2. Social preference reversals under harmful expansions of the outcome set

Independence of Harmless Expansions guarantees that the social ranking of two acts is unaffected by the addition of an outcome that all individuals find at least as good as the one they initially found worst. It does not prevent the social ranking to be affected by the addition of an outcome that some agent finds worse than all initially relevant outcomes –a *harmful* expansion.

Such social preference reversals do occur under the Nash SWFs. The following example illustrates the possible extent of the phenomenon.

Let $N = \{1, 2\}$ and $X = \{x, y\}$, so that an act $a \in A(X)$ is completely determined by the set $E_a := \{s \in S \mid a(s) = x\}$. Let R_1, R_2 be Savage preferences over $A(X)$ such that xP_1y, yP_2x, and $p^*(R_1) = p^*(R_2) = p$. Under the uniform Nash SWF, the social preference at $R_{\{1,2\}}$ is represented by a social utility function $U(\cdot, X, R_{\{1,2\}})$ that is the product of the $(0, 1)$-normalized utilities of the individuals. Since $U^*(x, X, R_1) = U^*(y, X, R_2) = 1$ and $U^*(y, X, R_1) = U^*(x, X, R_2) = 0$,

$$\begin{aligned} U(a, X, R_{\{1,2\}}) &= U^*(a, X, R_1)U^*(a, X, R_2) \\ &= p(E_a)(1 - p(E_a)) \end{aligned}$$

for all $a \in A(X)$. Social utility is *single-peaked* in $p(E_a)$, the probability that the act a yields outcome x, with peak at $p(E_a) = \frac{1}{2}$.

Let now $X' = \{x, y, z\}$ and let R'_1, R'_2 be Savage preferences over $A(X')$ such that

$$\begin{aligned} U^*(x, X', R_1) &= 1,\ U^*(y, X', R_1) = 1 - \varepsilon,\ U^*(z, X', R_1) = 0, \\ U^*(x, X', R_2) &= 0,\ U^*(y, X', R_2) = \varepsilon,\ U^*(z, X', R_2) = 1, \end{aligned}$$

where $0 < \varepsilon < \frac{1}{2}$, and $p^*(R'_1) = p^*(R'_2) = p$. Notice that the profile $R'_{\{1,2\}}$ coincides with $R_{\{1,2\}}$ on $A(X)$ and individual 1 finds the new outcome z worse than all previously relevant

outcomes. For every $a \in A(X)$ (a binary act with outcome in $X = \{x, y\}$), the social utility of a at $R'_{\{1,2\}}$ is

$$\begin{aligned} U(a, X', R'_{\{1,2\}}) &= U^*(a, X', R'_1)U^*(a, X', R'_2) \\ &= [p(E_a) + (1 - p(E_a))(1 - \varepsilon)]\,[(1 - p(E_a))\varepsilon] \\ &= \varepsilon(1 - \varepsilon) + \varepsilon(2\varepsilon - 1)p(E_a) - \varepsilon^2\,[p(E_a)]^2\,. \end{aligned}$$

Since $0 < \varepsilon < \frac{1}{2}$, social utility is now *decreasing* in the probability that a yields outcome x. In particular, the social ranking of all acts for which this probability is below $\frac{1}{2}$ is reversed.

It should be stressed that such reversals are not proper to the Nash SWFs. Indeed, an obvious corollary to our Theorem is that *no* SWF satisfies the Weak Pareto Principle, the full independence axiom obtained by dropping proviso (iii) from the premise of Independence of Harmless Expansions, and Continuity. This impossibility may be interpreted as a version of Arrow's theorem in the context of aggregation of Savage preferences.

However, it *is* possible to guarantee that society's preference is unaffected by a harmful expansion if the latter is also *useless*, i.e., if no individual finds the new outcome better than all the initially relevant ones. Specifically, consider the following axiom:

Independence of Useless Expansions. For all $(X, R_N), (X', R'_N) \in \mathcal{D}$, if (i) $X \subseteq X'$, (ii) R'_N coincides with R_N on $A(X)$, and (iii) for all $x' \in X'$ and $i \in N$ there exists some $x_i \in X$ such that $x_i R'_i x'$, then $\mathbf{R}(X', R'_N)$ coincides with $\mathbf{R}(X, R_N)$ on $A(X)$.

This requirement is exactly dual to Independence of Harmless Expansions. It makes sense in burden-sharing contexts where the best relevant outcome for each individual can easily be identified but the worst outcome cannot. Given our Theorem, it is straightforward to show that the Weak Pareto Principle, Independence of Useless Expansions, and Continuity characterize a class of SWFs that are dual to the belief-weighted Nash SWFs. Define the $(-1, 0)$-normalized representation of a preference $R \in \mathcal{R}^*(X)$ to be the Savage representation $U(\cdot, u, p)$ of R such that

$$\inf_X u = -1 \text{ and } \sup_X u = 0.$$

Denote it ${}^*U(\cdot, X, R)$. A SWF $\mathbf{R}$ is a *(belief-weighted) dual Nash SWF* if there is a function $\gamma : \mathcal{P}^N \to \Delta_N$ such that, for all $(X, R_N) \in \mathcal{D}$ and all $a, b \in A(X)$,

$$a\mathbf{R}(X, R_N)b \Leftrightarrow -\prod_{i \in N} (-\,{}^*U(a, X, R_i))^{\gamma_i(p^*(R_N))} \geq -\prod_{i \in N} (-\,{}^*U(b, X, R_i))^{\gamma_i(p^*(R_N))}$$

Contrary to the Nash SWFs, these dual Nash SWFs are unattractive from the point of view of fairness. Because the function $(w_1, ..., w_n) \mapsto -\prod_{i \in N}(-w)_i^{c_i}$ is strictly quasi-*convex* on $[-1, 0)^N$ (when $c \gg 0$), the corresponding SWF is "equality-averse". As an illustration, suppose $N = \{1, 2\}$, $X = \{x, y\}$, let $\mathbf{R}$ be the uniform dual Nash SWF, and let $R_{\{1,2\}}$ be a profile such that $p^*(R_1) = p^*(R_2) = p$ and ${}^*U(x, X, R_1) = {}^*U(y, X, R_2) = 0$ and ${}^*U(y, X, R_1) = {}^*U(x, X, R_2) = -1$. Then

$$-(-\,{}^*U(x, X, R_1))(-\,{}^*U(x, X, R_2)) = 0$$

whereas for any event E such that $0 < p(E) < 1$,

$$-(-\ ^*U(x_Ey, X, R_1))(-^*\ U(x_Ey, X, R_2)) = -(1 - p(E))(p(E)),$$

hence, $x\ \mathbf{P}(X, R_{\{1,2\}})x_Ey$. Society prefers x, which yields individual 1's favorite and 2's worst outcome in all states, to x_Ey, which gives each individual a chance to get her best outcome.

5.3. Indeterminacy –and how to reduce it

The weight function γ associated with a belief-weighted Nash SWF $\mathbf{R}$ is not constrained by the three axioms in the Theorem. Restrictions on γ may be obtained by imposing further axioms on $\mathbf{R}$. We explore here the consequences of Anonymity and State Neutrality. The former is the usual requirement that the social preference should not be affected by a relabeling of the individuals. The latter is a new axiom saying that the labeling of the *states of nature* should be irrelevant.

Some notation is needed to define these properties. Let $\Pi(N)$ be the set of bijections from N into itself and let $\Pi(S)$ be the set of bijections from S into itself. For any $X \in \mathcal{X}$, $\pi \in \Pi(N)$, $p_N \in \mathcal{P}^N$, $R_N \in \mathcal{R}(X)^N$, and $z \in \Delta_N$, define ${}^{\pi}p_N \in \mathcal{P}^N$, ${}^{\pi}R_N \in \mathcal{R}(X)^N$, and ${}^{\pi}z \in \Delta_N$ by ${}^{\pi}p_{\pi(i)} = p_i$, ${}^{\pi}R_{\pi(i)} = R_i$, and ${}^{\pi}z_{\pi(i)} = z_i$ for all $i \in N$.

Anonymity. For all $(X, R_N) \in \mathcal{D}$ and $\pi \in \Pi(N)$, $\mathbf{R}(X, {}^{\pi}R_N) = \mathbf{R}(X, R_N)$.

For any $X \in \mathcal{X}$, $\pi \in \Pi(S)$, $a \in A(X)$, $p \in \mathcal{P}$, and $R \in \mathcal{R}(X)$, define $\pi a \in A(X)$ by $(\pi a)(\pi(s)) = a(s)$ for all $s \in S$, define $\pi p \in \mathcal{P}$ by $(\pi p)(\pi(E)) = p(E)$ for all $E \subseteq S$, and define $\pi R \in \mathcal{R}(X)$ by $\pi a\ \pi R\ \pi b \Leftrightarrow aRb$ for all $a, b \in A(X)$. If $p_N \in \mathcal{P}^N$ and $R_N \in \mathcal{R}(X)^N$, let $\pi p_N = (\pi p_1, ..., \pi p_n)$ and $\pi R_N = (\pi R_1, ..., \pi R_n)$.

State Neutrality. For all $(X, R_N) \in \mathcal{D}$ and $\pi \in \Pi(S)$, $\mathbf{R}(X, \pi R_N) = \pi\mathbf{R}(X, R_N)$.

A more explicit formulation of the statement $\mathbf{R}(X, \pi R_N) = \pi\mathbf{R}(X, R_N)$ reads $a\mathbf{R}(X, R_N)b \Leftrightarrow \pi a\ \mathbf{R}(X, \pi R_N)\ \pi b$ for all $a, b \in A(X)$. This means that permuting states of nature yields a correspondingly permuted social ranking of the acts. Note that since constant acts are unchanged under any permutation of the states, State Neutrality implies that the social ranking of constant acts is unaffected by a relabeling of the states: $x\mathbf{R}(X, R_N)y \Leftrightarrow x\mathbf{R}(X, \pi R_N)y$ for all $x, y \in X$.

Call a weight function γ *symmetric* if $\gamma({}^{\pi}p_N) = {}^{\pi}\gamma(p_N)$ for all $p_N \in \mathcal{P}^N$ and all $\pi \in \Pi(N)$ and call it *invariant (under state relabeling)* if $\gamma(\pi p_N) = \gamma(p_N)$ for all $p_N \in \mathcal{P}^N$ and all $\pi \in \Pi(S)$.

Proposition. *If* $\mathbf{R}$ *is a belief-weighted Nash SWF with associated weight function* γ*, then*
(a) $\mathbf{R}$ *satisfies Anonymity if and only if* γ *is symmetric,*
(b) $\mathbf{R}$ *satisfies State Neutrality if and only if* γ *is invariant.*

Proof. Let $\mathbf{R}$ be a belief-weighted Nash SWF with associated weight function γ. The "if" part of statements (a) and (b) is easy to check.

Step 1. To prove the "only if" part of statement (a), let us assume that γ is not symmetric and show that $\mathbf{R}$ violates Anonymity. Since γ is not symmetric, there exist $p_N \in \mathcal{P}^N$, $\pi \in \Pi(N)$, and $i, j \in N$ such that

$$\gamma_i(p_N) < \gamma_{\pi(i)}({}^{\pi}p_N) \text{ and } \gamma_j(p_N) > \gamma_{\pi(j)}({}^{\pi}p_N).$$

Without loss of generality, suppose $i = 1$ and $j = 2$. Letting $c_1 := \gamma_1(p_N)$, $c_2 := \gamma_2(p_N)$, $c_1' := \gamma_{\pi(1)}({}^{\pi}p_N)$, and $c_2' := \gamma_{\pi(2)}({}^{\pi}p_N)$, we rewrite the above inequalities as

$$c_1 < c_1' \text{ and } c_2 > c_2'.$$

These inequalities imply that there exist numbers $k_1, k_2 \in (0,1)$ such that

$$\frac{c_1}{c_2} < \frac{k_2}{k_1} < \frac{c_1'}{c_2'}. \tag{5.4}$$

Fix two such numbers.

Let $X \in \mathcal{X}$ and let $x, y \in X$. For every $\varepsilon \in (0,1)$, define the real-valued functions $u_1^\varepsilon, u_2^\varepsilon, u_3, ..., u_n$ on X by

$$\begin{aligned}
\inf_X u_1^\varepsilon &= 0,\ u_1^\varepsilon(x) = \frac{1}{2}(1-\varepsilon k_1),\ u_1^\varepsilon(y) = \frac{1}{2},\ \sup_X u_1^\varepsilon = 1,\\
\inf_X u_2^\varepsilon &= 0,\ u_2^\varepsilon(x) = \frac{1}{2}(1+\varepsilon k_2),\ u_2^\varepsilon(y) = \frac{1}{2},\ \sup_X u_2^\varepsilon = 1,\\
\inf_X u_i &= 0,\ u_i(x) = u_i(y) = \sup_X u_i = 1 \text{ for } i = 3, ..., n.
\end{aligned}$$

For $i = 1, 2$, let $R_i^\varepsilon \in \mathcal{R}^*(X)$ be the preference with Savage representation $U(\cdot, u_i^\varepsilon, p_i)$ and, for $i = 3, ..., n$, let $R_i \in \mathcal{R}^*(X)$ be the preference with Savage representation $U(\cdot, u_i, p_i)$. Let $R_N^\varepsilon = (R_1^\varepsilon, R_2^\varepsilon, R_3, ..., R_n)$. We claim that

$$x\mathbf{P}(X, R_N^\varepsilon)y \text{ and } y\mathbf{P}(X,\ {}^{\sigma}R_N^\varepsilon)x \tag{5.5}$$

for ε small enough, meaning that $\mathbf{R}$ violates Anonymity.

To establish that (5.5) holds when ε is small enough, note that, because $U(x, u_i, p_i) = U(y, u_i, p_i) = 1$ for $i = 3, ..., n$, we have

$$\begin{aligned}
x\mathbf{P}(X, R_N^\varepsilon)y &\Leftrightarrow [U(x, u_1^\varepsilon, p_1)]^{c_1}\,[U(x, u_2^\varepsilon, p_2)]^{c_2} > [U(y, u_1^\varepsilon, p_1)]^{c_1}\,[U(y, u_2^\varepsilon, p_2)]^{c_2}\\
&\Leftrightarrow \left[\frac{1}{2}(1-\varepsilon k_1)\right]^{c_1}\left[\frac{1}{2}(1+\varepsilon k_2)\right]^{c_2} > \left[\frac{1}{2}\right]^{c_1}\left[\frac{1}{2}\right]^{c_2}\\
&\Leftrightarrow (1-\varepsilon k_1)^{c_1}(1+\varepsilon k_2)^{c_2} > 1
\end{aligned}$$

whereas

$$\begin{aligned}
y\mathbf{P}(X,\ {}^{\pi}R_N^\varepsilon)x &\Leftrightarrow [U(y, u_1^\varepsilon, p_1)]^{c_1'}\,[U(y, u_2^\varepsilon, p_2)]^{c_2'} > [U(x, u_1^\varepsilon, p_1)]^{c_1'}\,[U(x, u_2^\varepsilon, p_2)]^{c_2'}\\
&\Leftrightarrow 1 > (1-\varepsilon k_1)^{c_1'}(1+\varepsilon k_2)^{c_2'}.
\end{aligned}$$

It remains to check that $(1-\varepsilon k_1)^{c_1}(1+\varepsilon k_2)^{c_2} > 1 > (1-\varepsilon k_1)^{c_1'}(1+\varepsilon k_2)^{c_2'}$ when ε is small. Define the real-valued functions W, W' on $\mathbb{R}^2_+$ by $W(z_1,z_2) = z_1^{c_1}z_2^{c_2}$, $W'(z_1,z_2) = z_1^{c_1'}z_2^{c_2'}$. Observe that, when ε is small, $(1-\varepsilon k_1)^{c_1}(1+\varepsilon k_2)^{c_2} = W(1-\varepsilon k_1, 1+\varepsilon k_2) > 1$ if $W(1,1) - \frac{\partial W}{\partial z_1}(1,1)\varepsilon k_1 + \frac{\partial W}{\partial z_2}(1,1)\varepsilon k_2 > 1$. The latter inequality holds if and only if $c_2k_2 - c_1k_1 > 0$, which is guaranteed by the first inequality in (5.4). Likewise, when ε is small, $(1-\varepsilon k_1)^{c_1'}(1+\varepsilon k_2)^{c_2'} < 1$ if $c_2'k_2 - c_1'k_1 < 0$, which holds because of the second inequality in (5.4).

Step 2. To prove the "only if" part of statement (b), let us assume that γ is not invariant and show that $\mathbf{R}$ violates State Neutrality. Since γ is not invariant, there exist $p_N \in \mathcal{P}^N$, $\pi \in \Pi(S)$, and $i, j \in N$ such that

$$\gamma_i(p_N) < \gamma_i(\pi p_N) \text{ and } \gamma_j(p_N) > \gamma_j(\pi p_N).$$

Suppose again that $i = 1$ and $j = 2$ and let now $c_1 := \gamma_1(p_N)$, $c_2 := \gamma_2(p_N)$, $c_1' := \gamma_1(\pi p_N)$, and $c_2' := \gamma_2(\pi p_N)$, so that $c_1 < c_1'$ and $c_2 > c_2'$, guaranteeing that there exist numbers $k_1, k_2 \in (0,1)$ satisfying (5.4). Fix two such numbers.

Fix again $X \in \mathcal{X}$, $x, y \in X$, $\varepsilon \in (0,1)$, and consider the profile $R_N^\varepsilon = (R_1^\varepsilon, R_2^\varepsilon, R_3, ..., R_n)$ constructed in Step 1. As we have shown,

$$x\mathbf{P}(X, R_N^\varepsilon)y \;\Leftrightarrow (1-\varepsilon k_1)^{c_1}(1+\varepsilon k_2)^{c_2} > 1.$$

Next, because $U(x, u_i, \pi p_i) = U(x, u_i, p_i) = 1$ and $U(y, u_i, \pi p_i) = U(y, u_i, p_i) = 1$ for $i = 3, ..., n$,

$$\begin{aligned}
y\mathbf{P}(X, \pi R_N^\varepsilon)x \quad &\Leftrightarrow \quad \left[U(y, u_1^\varepsilon, \pi p_1)\right]^{c_1'}\left[U(y, u_2^\varepsilon, \pi p_2)\right]^{c_2'} > \left[U(x, u_1^\varepsilon, \pi p_1)\right]^{c_1'}\left[U(x, u_2^\varepsilon, \pi p_2)\right]^{c_2'} \\
&\Leftrightarrow \quad \left[u_1^\varepsilon(y)\right]^{c_1'}\left[u_2^\varepsilon(y)\right]^{c_2'} > \left[u_1^\varepsilon(x)\right]^{c_1'}\left[u_2^\varepsilon(x)\right]^{c_2'} \\
&\Leftrightarrow \quad \left[\frac{1}{2}\right]^{c_1'}\left[\frac{1}{2}\right]^{c_2'} > \left[\frac{1}{2}(1-\varepsilon k_1)\right]^{c_1'}\left[\frac{1}{2}(1+\varepsilon k_2)\right]^{c_2'} \\
&\Leftrightarrow \quad 1 > (1-\varepsilon k_1)^{c_1'}(1+\varepsilon k_2)^{c_2'}.
\end{aligned}$$

Since we have shown in Step 1 that $(1-\varepsilon k_1)^{c_1}(1+\varepsilon k_2)^{c_2} > 1 > (1-\varepsilon k_1)^{c_1'}(1+\varepsilon k_2)^{c_2'}$ when ε is small, we conclude that $x\mathbf{P}(X, R_N^\varepsilon)y$ and $y\mathbf{P}(X, \pi R_N^\varepsilon)x$ for ε small enough, meaning that $\mathbf{R}$ violates State Neutrality.■

We have not found a compact characterization of the weight functions γ that are both symmetric and invariant. The following example, however, shows that the two conditions jointly do *not* force $\gamma(p_N) = (\frac{1}{n}, ..., \frac{1}{n})$ for every $p_N \in \mathcal{P}^N$. For each $i \in N$ and $p_N \in \mathcal{P}^N$, let $N(i, p_N) = \{j \in N \mid p_j = p_i\}$ and define

$$\gamma_i(p_N) = \frac{|N(i, p_N)|}{\sum_{j \in N} |N(j, p_N)|}.$$

Under this function γ, the weight attached to an individual's utility is proportional to the number of individuals sharing her belief. We do not claim that the resulting Nash SWF is appealing, and it is clear that symmetric and invariant weight functions which favor eccentric (rather than popular) beliefs can also be constructed. What the example does show is that a defense of the uniform Nash SWF (corresponding to $\gamma(p_N) = (\frac{1}{n}, ..., \frac{1}{n})$ for every $p_N \in \mathcal{P}^N$) requires going beyond traditional symmetry and invariance requirements.

6. References

Arrow, K. (1963). *Social choice and individual values,* 2nd edition, New York: Wiley.

Baccelli, J. (2017). "Do bets reveal beliefs? A unified perspective on state-dependent utility issues," *Synthese* **194**, 3393-3419.

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

Börgers, T. and Choo, Y.-M. (2017b). "Revealed relative utilitarianism," Mimeo, University of Michigan.

Chambers, C. and Hayashi, T. (2006). "Preference aggregation under uncertainty: Savage vs. Pareto," *Games and Economic Behavior* **54**, 430-440.

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

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

Diamond, P. (1967). "Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment," *Journal of Political Economy* **75**, 765-766.

Duffie, D. (2014). "Challenges to a policy treatment of speculation trading motivated by differences in beliefs," *Journal of Legal Studies* **43**, 173-182.

Gilboa, I., Samet, D., and Schmeidler, D. (2004). "Utilitarian aggregation of beliefs and tastes," *Journal of Political Economy* **112**, 932-938.

Gilboa, I., Samuelson, L., and Schmeidler, D. (2014), "No-betting Pareto dominance," *Econometrica* **82**, 1405–1442.

Harsanyi, J. (1955). "Cardinal welfare, individualistic ethics, and interpersonal comparison of utility," *Journal of Political Economy* **63**, 309-332.

Hylland, A. and Zeckhauser, R. (1979), "The impossibility of Baeysian group decision making with separate aggregation of beliefs and values," *Econometrica* **47**, 1321–1336.

Kaneko, M. and Nakamura, K. (1979). "The Nash social welfare function," *Econometrica* **47**, 423-435.

Karni, E. (1998). "Impartiality: definition and representation," *Econometrica* **66**, 1405-1415.

Mongin, P. (1995). "Consistent Bayesian aggregation," *Journal of Economic Theory* **66**, 313-351.

Mongin, P. (1997). "Spurious unanimity and the Pareto principle," THEMA Working Paper, Université de Cergy-Pontoise.

Mongin, P. (1998). "The paradox of the Bayesian experts and state-dependent utility theory," *Journal of Mathematical Economics* **29**, 331-361.

Mongin, P. and Pivato, M. (2016). "Social preferences under twofold uncertainty," HEC Paris Research Paper ECO/SCD-2016-1154.

Nash, J. (1950). "The bargaining problem," *Econometrica* **18**, 155-162.

Osborne, D.K. (1976), "Irrelevant alternatives and social welfare," *Econometrica* **44**, 1001-1015.

Savage, L. and Aumann, R. (1987). "Letters between Leonard Savage and Robert Aumann (1971)," in J. Drèze (Ed.), *Essays on economic decisions under uncertainty*, 76–81. Cambridge: Cambridge University Press.

Segal, U. (2000). "Let's agree that all dictatorships are equally bad," *Journal of Political Economy* **108**, 569-589.

Sen, A. (1970). *Collective choice and social welfare,* San Francisco: Holden-Day.

West, M. (1984). "Bayesian aggregation," *Journal of the Royal Statistical Society* (Series A) **147**, 600-607.