April 15, 2010

Generalized Utilitarianism and Harsanyi's Impartial Observer Theorem*

Abstract

Harsanyi's impartial observer must consider two types of lotteries: imaginary identity lotteries ("accidents of birth") which she faces as herself, and the real outcome lotteries ("life chances") to be faced by the individuals she imagines becoming. If we maintain a distinction between identity and outcome lotteries then Harsanyi-like axioms yield generalized utilitarianism, and allow us to accommodate concerns about different individuals' risk attitudes and concerns about fairness. Requiring an impartial observer to be indifferent as to which individual should face similar risks restricts her social welfare function, but still allows her to accommodate fairness. Requiring an impartial observer to be indifferent between identity and outcome lotteries, however, forces her to ignore both fairness and different risk-attitudes, and yields a new axiomatization of Harsanyi's utilitarianism.

Keywords: generalized utilitarianism, impartial observer, social welfare function, fairness, ex ante egalitarianism.

JEL Classification: D63, D71

Simon Grant
Department of Economics
Rice University

Atsushi Kajii
Insitute of Economic Research
Kyoto University

Ben Polak
Department of Economics & School of Management
Yale University

Zvi Safra
University of Exeter, The College of Management and Tel Aviv University

*We thank John Broome, Jurgen Eichberger, Marc Fleurbaey, Edi Karni, Bart Lipman, Philippe Mongin, Stephen Morris, Heve Moulin, Klaus Nehring, David Pearce, John Quiggin, John Roemer, John Weymark, three referees and the editor for many helpful comments. Atsushi Kajii thanks Grant-in-Aid for Scientific Research S (grant no. 90152298) and the Inamori foundation for their support. Zvi Safra thanks the Israel Science Foundation (grant no. 1299/05) and the Henry Crown Institute of Business Research for their support.

1 Introduction

This paper revisits Harsanyi's (1953, 1955, 1977) utilitarian impartial observer theorem. Consider a society of individuals $\mathcal{I}$. The society has to choose among different social policies, each of which induces a probability distribution or 'lottery' ℓ over a set of social outcomes $\mathcal{X}$. Each individual i has preferences $\succsim_i$ over these lotteries. These preferences are known, and they differ.

To help choose among social policies, Harsanyi proposed that each individual should imagine herself as an 'impartial observer' who does not know which person she will be. That is, the impartial observer faces not only the real lottery ℓ over the social outcomes in $\mathcal{X}$, but also a hypothetical lottery z over which identity in $\mathcal{I}$ she will assume. In forming preferences $\succsim$ over all such 'extended lotteries', an impartial observer is forced to make interpersonal comparisons: for example, she is forced to compare being person i in social state x with being person j in social state x'.

Harsanyi assumed the so-called 'acceptance principle'; that is, when an impartial observer imagines herself being person i she adopts person i's preferences over the outcome lotteries. He also assumed that all individuals are expected utility maximizers, and that they continue to be so in the role of the impartial observer. Harsanyi argued that these "Bayesian rationality" axioms force the impartial observer to be a (weighted) utilitarian. More formally, over all extended lotteries (z, ℓ) in which the identity lottery and the outcome lotteries are independently distributed, the impartial observer's preferences admit a representation of the form

$$V(z, \ell) = \sum\nolimits_i z_i U_i(\ell) \tag{1}$$

where z_i is the probability of assuming person i's identity and $U_i(\ell) := \int_{\mathcal{X}} u_i(x)\, \ell(dx)$ is person i's von Neumann-Morgenstern expected utility for the outcome lottery ℓ. Where no confusion arises, we will omit the "weighted" and refer to the representation in (1) simply as utilitarianism.[1]

[1] Some writers (e.g., Sen 1970, 1977; Weymark 1991, Mongin 2001, 2002) reserve the term utilitarianism for social welfare functions in which all the z_i's are equal and the U_i's are welfares not just von-Neumann Morgenstern utilities. Harsanyi claims that impartial observers should assess social policies using equal z_i weights, and that von-Neumann Morgenstern utilities should be identified with welfares. Harsanyi (1977, pp. 57-60) concedes that his axioms do not force all potential impartial observers to agree in their extended preferences. Nevertheless, he claims that, given enough information about "the individuals' psychological, biological and cultural characteristics" all impartial observers would agree. These extra claims are not the focus of this paper, but we will return to the issues of agreement and welfare in section 7.

Harsanyi's utilitarianism has attracted many criticisms. We confront just two: one concerning fairness; and one concerning different attitudes toward risk. To illustrate both criticisms, consider two individuals, i and j and two social outcomes x_i and x_j. Person i strictly prefers outcome x_i to outcome x_j, but person j strictly prefers x_j to x_i. Perhaps, there is some (possibly indivisible) good, and x_i is the state in which person i gets the good while x_j is the state in which person j gets it. Suppose that an impartial observer would be indifferent between being person i in state x_i and being person j in state x_j; hence $u_i(x_i) = u_j(x_j) =: u^H$. She is also indifferent between being i in x_j and being j in x_i; hence $u_i(x_j) = u_j(x_i) =: u^L$. And she strictly prefers the first pair (having the good) to the second (not having the good); hence $u^H > u^L$.

The concern about fairness is similar to Diamond's (1967) critique of Harsanyi's aggregation theorem. Consider the two extended lotteries illustrated in tables (a) and (b) in which rows are the people and columns are the outcomes.

	x_i	x_j
i	1/2	0
j	1/2	0

(a)

	x_i	x_j
i	1/4	1/4
j	1/4	1/4

(b)

In each, the impartial observer has a half chance of being person i or person j. But in table (a), the good is simply given outright to person i: outcome x_i has probability 1. In table (b), the good is allocated by tossing a coin: the outcomes x_i and x_j each have probability 1/2. Diamond argued that a fair-minded person might prefer the second allocation policy since it gives each person a "fair shake".[2] But Harsanyi's utilitarian impartial observer is indifferent to such considerations of fairness. Each policy (or its associated extended lottery) involves a half chance of getting the good and hence yields the impartial observer $\frac{1}{2}u^H + \frac{1}{2}u^L$. The impartial observer cares only about her total chance of getting the good, not how this chance is distributed between person i and j.

The concern about different risk attitudes is less familiar.[3] Consider the two extended lotteries

[2] Societies often use both simple lotteries and weighted lotteries to allocate goods (and bads), presumably for fairness considerations. Examples include the draft, kidney machines, oversubscribed events, schools, and public housing, and even whom should be thrown out of a lifeboat! For a long list and an enlightening discussion, see Elster (1989).

[3] Pattanaik (1968) remarks that in reducing an identity-outcome lottery to a one-stage lottery, "what we are

illustrated in tables (c) and (d).

	x_i	x_j
i	1/2	1/2
j	0	0

(c)

	x_i	x_j
i	0	0
j	1/2	1/2

(d)

In each, the impartial observer has a half chance of being in state x_i or state x_j, and hence a half chance of getting the good. But in (c), the impartial observer faces this risk as person i, while in (d), she faces the risk as person j. Suppose that person i is more comfortable facing such a risk than is person j.[4] But Harsanyi's utilitarian impartial observer is indifferent to such considerations of risk attitude. Each of the extended lotteries (c) and (d) again yield $\frac{1}{2}u^H + \frac{1}{2}u^L$. Thus, Harsanyi's impartial observer does not care who faces this risk.

In his own response to the concern about fairness, Harsanyi (1975) argued that, even if randomizations were of value for promoting fairness (which he doubted), any explicit randomization is superfluous since 'the great lottery of (pre-)life' may be viewed as having already given each child an equal chance of being each individual. That is, for Harsanyi, it does not matter whether a good is allocated by a (possibly imaginary) lottery over identities as in table (a) above, or by a (real) lottery over outcomes as in table (c), or by some combination of the two as in table (b). The dispute about fairness thus seems to rest on whether or not we are indeed indifferent between identity and outcome lotteries; that is, between 'accidents of birth' and real 'life chances'. For Harsanyi, they are equivalent, but, for those concerned about fairness, 'genuine' life chances might be preferred to 'mere' accidents of birth.[5]

If we regard outcome and identity lotteries as equivalent, there is little scope left to accommodate different risk attitudes of different individuals. For example, the outcome lottery in table (c)

actually doing is to combine attitudes to risk of more than one person" (pp. 1165-6).

[4] To make this notion of greater 'comfort' concrete, suppose that both people have certainty equivalents for the risk of a half chance of being in states x_i or x_j – call these certainty equivalents y_i and y_j respectively – and suppose that, according to the interpersonal comparisons of the impartial observer, person j is prepared to give up more than person i to remove this risk: that is, the impartial observer would prefer to be person i with y_i than person j with y_j. In this case, by the definition of a certainty equivalent, the acceptance principle and transitivity, the impartial observer would prefer to face the risk of a half chance of being in states x_i or x_j as person i than as person j.

[5] This could be seen as an example of what Ergin & Gul (2009) call issue or source preference.

would be indifferent to the identity lottery in table (a) even though the risk in the first is faced by person i and the risk in the second is faced by the impartial observer. Similarly, the outcome lottery in (d) would be indifferent to the identity lottery in (a). Hence the two outcome lotteries (c) and (d) must be indifferent even though one is faced by person i and the other by person j. In effect, indifference between outcome and identity lotteries treats all risks as if they were faced by one agent, the impartial observer: it forces us to conflate the risk attitudes of individuals with those of the impartial observer herself. But Harsanyi's own acceptance principle states that, when the impartial observer imagines herself as person i, she should adopt person i's preferences over the outcome lotteries faced by person i. This suggests that different lotteries perhaps should not be treated as equivalent if they are faced by different people with possibly different risk attitudes.

We want to make explicit the possibility that an impartial observer might distinguish between the identity lotteries $\triangle(\mathcal{I})$ she faces and the outcome lotteries $\triangle(\mathcal{X})$ faced by the indviduals. Harsanyi's impartial observer is assumed to form preferences over the entire set of joint distributions $\triangle(\mathcal{I} \times \mathcal{X})$ over identities and outcomes. In such a set up, it is hard to distinguish outcome from identity lotteries since the resolution of identity can partially or fully resolve the outcome. For example, the impartial observer could face a joint distribution in which, if she becomes person i then society holds the outcome lottery ℓ, but if she becomes person j then social outcome x obtains for sure. To keep this distinction clean, we restrict attention to product lotteries, $\triangle(\mathcal{I}) \times \triangle(\mathcal{X})$. That is, the impartial observer only forms preferences over extended lotteries in which the outcome lottery she faces is the same regardless of which identity she assumes. That said, our restriction to product lotteries is for conceptual clarity only and is not essential for the main results.[6]

Harsanyi's assumption that identity and outcome lotteries are equivalent is implicit. Suppose that, without imposing such an equivalence, we impose each of Harsanyi's three main assumptions: that, if the impartial observer imagines being individual i, she *accepts* the preferences of that individual; that each individual satisfies *independence* over the lotteries he faces (which are outcome lotteries); and that the impartial observer satisfies *independence* over the lotteries she

[6] See section 6 below.

faces (which are identity lotteries). Notice that, by acceptance, the impartial observer inherits independence over outcome lotteries. But this is not enough to force us to the (weighted) utilitarianism of expression 1. Instead (theorem 1), we obtain a generalized (weighted) utilitarian representation:

$$V(z,\ell) = \sum_i z_i \phi_i \left(U_i(\ell)\right) \tag{2}$$

where z_i is again the probability of assuming person i's identity and $U_i(\ell)$ is again person i's expected utility from the outcome lottery ℓ, but each $\phi_i(.)$ is a (possibly non-linear) transformation of person i's expected utility. Generalized utilitarianism is well known to welfare economists, but has not before been given foundations in the impartial-observer framework.[7]

Generalized utilitarianism can accommodate concerns about fairness if the ϕ_i-functions are concave.[8] Harsanyi's utilitarianism can be thought of as the special case where each ϕ_i is affine. The discussion above suggests that these differences about fairness involve preferences between identity and outcome lotteries. The framework allows us to formalize this intuition: we show that a generalized utilitarian impartial observer has concave ϕ_i-functions if and only if she has a preference for outcome lotteries over identity lotteries (i.e., a 'preference for life chances'); and she is a utilitarian if and only if she is indifferent between outcome and identity lotteries (i.e., 'indifferent between life chances and accidents of birth').[9]

Generalized utilitarianism can accommodate concerns about different risk attitudes simply by allowing the ϕ_i-functions to differ in their degree of concavity or convexity.[10] In the example above, the impartial observer first assessed equal welfares to being person i in state x_i or person j in state x_j, and equal welfares to being i in x_j or j in x_i. The issue of different risk attitudes seemed to

[7] For example, see Blackorby, Bossert and Donaldson (2005, chapter 4) and Blackorby, Donaldson & Mongin (2004). Both obtain similar representations for aggregating utility vectors; the former from Gorman-like separability assumptions, the latter by assuming consistency between evaluations based on the ex post social welfares and those based on ex ante utilities. See also Blackorby, Donaldson & Weymark (1999).

[8] In our story, we have $\phi_i(u_i(x_i)) = \phi_j(u_j(x_j)) > \phi_i(u_i(x_j)) = \phi_j(u_j(x_i))$. Thus, if the ϕ-functions are (strictly) concave, the impartial observer evaluatation of allocation policy (c) $\phi_i\left(\frac{1}{2}u_i(x_i) + \frac{1}{2}u_i(x_j)\right) > \frac{1}{2}\phi_i(u_i(x_i)) + \frac{1}{2}\phi_i(u_i(x_j)) = \frac{1}{2}\phi_i(u_i(x_i)) + \frac{1}{2}\phi_j(u_j(x_i))$, her evaluation of policy (a). The argument comparing (b) and (a) is similar.

[9] This provides a new axiomatization of Harsanyi's utilitarianism, distinct from, for example, Karni & Weymark (1998) or Safra & Weissengrin (2003).

[10] For example, if ϕ_i is strictly concave but ϕ_j is linear, then the impartial observer's evaluation of policy (c) $\phi_i\left(\frac{1}{2}u_i(x_i) + \frac{1}{2}u_i(x_j)\right) > \frac{1}{2}\phi_i(u_i(x_i)) + \frac{1}{2}\phi_i(u_i(x_j)) = \frac{1}{2}\phi_j(u_j(x_j)) + \frac{1}{2}\phi_j(u_j(x_i)) = \phi_j\left(\frac{1}{2}u_j(x_j) + \frac{1}{2}u_j(x_i)\right)$, her evaluation of policy (d).

rest on whether such 'equal welfares' implies equal von-Neumann Morgenstern utilities. We show that a generalized utilitarian impartial observer uses the same ϕ-function for all people (implying the same mapping from their von-Neumann Morgenstern utilities to her welfare assessments) if and only if she would be indifferent as to which person to be when facing such similar risks.

Where does Harsanyi implicitly assume both indifference between life chances and accidents of births and indifference between individuals facing similar risks? Harsanyi's independence axiom goes further than ours in two ways. First, in our case, the impartial observer inherits independence over outcome lotteries indirectly (via acceptance) from individuals' preferences. In contrast, Harsanyi's axiom imposes independence over outcome lotteries directly on the impartial observer. We will see that this direct imposition forces the impartial observer to be indifferent as to which individual faces similar risks. Second, Harsanyi's independence axiom extends to randomizations that simultaneously mix outcome and identity lotteries. We will see that this assumption forces the impartial observer to be indifferent between these two types of randomization, and this in turn precludes concern for fairness.

Earlier attempts to accommodate fairness considerations focussed on dropping independence. For example, Karni & Safra (2002) relax independence for the individual preferences, while Epstein & Segal (1992) relax independence for the impartial observer.[11] Our approach maintains independence for each agent but restricts its domain to the lotteries faced by that agent.

Section 2 sets up the framework. Section 3 axiomatizes generalized utilitarianism. Section 4 deals with concerns about fairness. We show that the impartial observer ignoring these concerns is equivalent to her being indifferent between identity and outcome lotteries. This yields a new axiomatization of Harsanyi's utilitarianism. Section 5 deals with concerns about different risk attitudes. Section 6 first shows how to extend our analysis to the entire set of joint distributions $\triangle(\mathcal{I} \times \mathcal{X})$ over identities and outcomes. We then show how Harsanyi's independence axiom restricted to our domain of product lotteries, $\triangle(\mathcal{I}) \times \triangle(\mathcal{X})$, implies both our independence axiom and both our indifference conditions: indifference between outcome and identity lotteries,

[11] Strictly speaking, Epstein & Segal's paper is in the context of Harsanyi's (1955) aggregation theorem. In addition, Broome (1991) addresses fairness concerns by expanding the outcome space to include the means of allocation (e.g., the use of a physical randomization device) as part of the description of the final outcome.

and indifference as to whom faces similar risks. Section 7 considers four possible views (including the one taken in this paper) for the role of the impartial observer. For each view we ask: what are the knowledge requirements for the impartial observer; and must all potential impartial observers agree in their preferences over extended lotteries; and we relate these to the issues of fairness and different risk attitudes. Proofs are in the appendix. Appendix B [on line] contains supplementary examples and proofs.

2 Set up and Notation

Let society consist of a finite set of individuals $\mathcal{I} = \{1, \ldots, I\}$, $I \geq 2$, with generic elements i and j. The set of social outcomes is denoted by $\mathcal{X}$ with generic element x. The set $\mathcal{X}$ is assumed to have more than one element and to be a compact metrizable space and associated with it is the set of events $\mathcal{E}$, which is taken to be the Borel sigma-algebra of $\mathcal{X}$. Let $\triangle(\mathcal{X})$ (with generic element ℓ) denote the set of *outcome lotteries*; that is the set of probability measures on $(\mathcal{X}, \mathcal{E})$ endowed with the weak convergence topology. These lotteries represent the risks actually faced by each individual in their lives. With slight abuse of notation, we will let x or sometimes $[x]$ denote the degenerate outcome lottery that assigns probability weight 1 to social state x.

Each individual i in $\mathcal{I}$, is endowed with a preference relation $\succsim_i$ defined over the set of life-chances $\triangle(\mathcal{X})$. We assume throughout that for each i in $\mathcal{I}$, the preference relation $\succsim_i$ is a complete, transitive binary relation on $\triangle(\mathcal{X})$, and that its asymmetric part $\succ_i$ is non-empty. We assume these preferences are continuous in that weak upper and weak lower contour sets are closed. Hence for each $\succsim_i$ there exists a non-constant function $V_i : \triangle(\mathcal{X}) \to \mathbb{R}$, satisfying for any ℓ and ℓ' in $\triangle(\mathcal{X})$, $V_i(\ell) \geq V_i(\ell')$ if and only if $\ell \succsim_i \ell'$. In summary, a society may be characterized by the tuple $\langle \mathcal{X}, \mathcal{I}, \{\succsim_i\}_{i \in \mathcal{I}} \rangle$.

In Harsanyi's story, the impartial observer imagines herself behind a veil of ignorance, uncertain about which identity she will assume in the given society. Let $\triangle(\mathcal{I})$ denote the set of *identity lotteries* on I. Let z denote the typical element of $\triangle(\mathcal{I})$, and let z_i denote the probability assigned by the identity lottery z to individual i. These lotteries represent the imaginary risks in the mind of the impartial observer of being born as someone else. With slight abuse of notation, we will let

i or sometimes $[i]$ denote the degenerate identity lottery that assigns probability weight 1 to the impartial observer assuming the identity of individual i.

As discussed above, we assume that the outcome and identity lotteries faced by the impartial observer are independently distributed; that is, she faces a product lottery $(z, \ell) \in \triangle(\mathcal{I}) \times \triangle(\mathcal{X})$. We shall sometimes refer to this as a product *identity-outcome lottery* or, where no confusion arises, simply as a product lottery.

Fix an impartial observer endowed with a preference relation $\succsim$ defined over $\triangle(\mathcal{I}) \times \triangle(\mathcal{X})$. We assume throughout that $\succsim$ is complete, transitive continuous (in that weak upper and weak lower contour sets are closed in the product topology), and that its asymmetric part $\succ$ is non-empty, and so it admits a (non-trivial) continuous representation $V : \triangle(\mathcal{I}) \times \triangle(\mathcal{X}) \to \mathbb{R}$. That is, for any pair of product lotteries, (z, ℓ) and (z', ℓ'), $(z, \ell) \succsim (z', \ell')$ if and only if $V(z, \ell) \geq V(z', \ell')$.

Utilitarianism *We say that the impartial observer is a* (weighted) utilitarian *if her preferences* $\succsim$ *admit a representation* $\left\langle \{U_i\}_{i \in \mathcal{I}} \right\rangle$ *of the form*

$$V(z, \ell) = \sum_{i=1}^{I} z_i U_i(\ell)$$

where, for each individual i *in* I, $U_i : \triangle(\mathcal{X}) \to \mathbb{R}$ *is a von Neumann-Morgenstern expected-utility representation of* $\succsim_i$*; i.e.,* $U_i(\ell) := \int_{\mathcal{X}} u_i(x)\, \ell(dx)$.

Generalized Utilitarianism *We say that the impartial observer is a* generalized (weighted) utilitarian *if her preferences* $\succsim$ *admit a representation* $\left\langle \{U_i, \phi_i\}_{i \in \mathcal{I}} \right\rangle$ *of the form*

$$V(z, \ell) = \sum_{i=1}^{I} z_i \phi_i \left[U_i(\ell) \right].$$

where, for each individual i *in* I, $\phi_i : \mathbb{R} \to \mathbb{R}$ *is a continuous, increasing function, and* $U_i : \triangle(\mathcal{X}) \to \mathbb{R}$ *is a von Neumann-Morgenstern expected-utility representation of* $\succsim_i$.

3 Generalized Utilitarianism

In this section, we axiomatize generalized utiltiarianism. The first axiom is Harsanyi's acceptance principle. In degenerate product lotteries of the form (i, ℓ) or (i, ℓ'), the impartial observer knows

she will assume identity i for sure. The acceptance principle requires that, in this case, the impartial observer's preferences $\succsim$ must coincide with that individual's preferences $\succsim_i$ over outcome lotteries.

Acceptance Principle. *For all i in $\mathcal{I}$ and all $\ell, \ell' \in \triangle(\mathcal{X})$, $\ell \succsim_i \ell'$ if and only if $(i, \ell) \succsim (i, \ell')$.*

Second, we assume that each individual i's preferences satisfy the independence axiom for the lotteries he faces; i.e., outcome lotteries.

Independence over Outcome Lotteries (for Individual i). *Suppose ℓ, $\ell' \in \triangle(\mathcal{X})$ are such that $\ell \sim_i \ell'$. Then, for all $\tilde{\ell}$, $\tilde{\ell}' \in \triangle(\mathcal{X})$, $\tilde{\ell} \succsim_i \tilde{\ell}'$ if and only if $\alpha\tilde{\ell}+(1-\alpha)\,\ell \succsim_i \alpha\tilde{\ell}'+(1-\alpha)\,\ell'$ for all α in $(0, 1]$.*

Third, we assume that the impartial observer's preferences satisfy independence for the lotteries she faces; i.e., identity lotteries. Here, however, we need to be careful. The set of product lotteries $\triangle(\mathcal{I}) \times \triangle(\mathcal{X})$ is not a convex subset of $\triangle(\mathcal{I} \times \mathcal{X})$ and hence not all probability mixtures of product lotteries are well defined. Thus, we adopt the following notion of independence.[12]

Independence over Identity Lotteries (for the Impartial Observer). *Suppose (z, ℓ), $(z', \ell') \in \triangle(\mathcal{I}) \times \triangle(\mathcal{X})$ are such that $(z, \ell) \sim (z', \ell')$. Then, for all $\tilde{z}$, $\tilde{z}' \in \triangle(\mathcal{I})$:$(\tilde{z}, \ell) \succsim (\tilde{z}', \ell')$ if and only if $(\alpha\tilde{z} + (1-\alpha)\,z, \ell) \succsim (\alpha\tilde{z}' + (1-\alpha)\,z', \ell')$ for all α in $(0, 1]$.*

To understand this axiom, first notice that the two mixtures on the right side of the implication are identical to $\alpha(\tilde{z}, \ell)+(1-\alpha)\,(z, \ell)$ and $\alpha(\tilde{z}', \ell')+(1-\alpha)\,(z', \ell')$ respectively. These two mixtures of product lotteries are well defined: they mix identity lotteries holding the outcome lottery fixed. Second, notice that the two product lotteries, (z, ℓ) and (z', ℓ'), that are 'mixed in' with weight $(1-\alpha)$ are themselves indifferent. The axiom states that 'mixing in' two indifferent lotteries (with equal weight) preserves the the original preference between $(\tilde{z}, \ell)$ and $(\tilde{z}', \ell')$ prior to mixing. Finally, notice that this axiom only applies to mixtures of identity lotteries holding the outcome

[12] This axiom is based on Fishburn's (1982, p. 88) and Safra & Weissengrin's (2003) substitution axioms for product lottery spaces. Their axioms, however, apply wherever probability mixtures are well defined in this space. We only allow mixtures of identity lotteries. In this respect, our axiom is similar to Karni & Safra's (2000) 'constrained independence' axiom, but their axiom applies to all joint distributions over identities and outcomes, not just to product lotteries.

lotteries fixed, not to the opposite case: mixtures of outcome lotteries holding the identity lotteries fixed.

To obtain our representation results, we work with a richness condition on the domain of individual preferences: we assume that none of the outcome lotteries under consideration are Pareto dominated.

Absence of Unanimity *For all* ℓ, $\ell' \in \triangle(\mathcal{X})$ *if* $\ell \succ_i \ell'$ *for some* i *in* $\mathcal{I}$ *then there exists* j *in* $\mathcal{I}$ *such that* $\ell' \succ_j \ell$.

This condition is perhaps a natural restriction in the context of Harsanyi's thought experiment. That exercise is motivated by the need to make social choices when agents disagree. We do not need to imagine ourselves as an impartial observer facing a identity lottery to rule out social alternatives that are Pareto dominated.[13]

These axioms are enough to yield a generalized utilitarian representation.

Theorem 1 (Generalized Utilitarianism) *Suppose that absence of unanimity applies. Then the impartial observer's preferences* $\succsim$ *admit a generalized utilitarian representation* $\left\langle \{U_i, \phi_i\}_{i \in \mathcal{I}} \right\rangle$ *if and only if the impartial observer satisfies the acceptance principle and independence over identity lotteries, and each individual satisfies independence over outcome lotteries.*

Moreover the functions U_i *are unique up to positive affine transformations and the composite functions* $\phi_i \circ U_i$ *are unique up to a common positive affine transformation.*

Grant et al. (2006: theorem 8) show that without absence of unanimity, we still obtain a generalized utilitarian representation but we lose the uniqueness of the composite functions $\phi_i \circ U_i$. Notice that, while the representation of each individual's preferences U_i is affine in outcome lotteries, in general, the representation of the impartial observer's preferences V is not.

[13] In Harsanyi's thought experiment, Pareto dominated lotteries would never be chosen by the impartial observer since the combination of the acceptance principle and Harsanyi's stronger independence axioms imply the Pareto criterion. We are grateful to a referee for this point.

4 Fairness or ex ante egalitarianism

So far we have placed no restriction on the shape of the ϕ_i-functions except that they are increasing. In a standard utilitarian social welfare function, each u_i-function maps individual i's income to an individual utility. These incomes differ across people, and concavity of the u_i-functions is associated with egalitarianism over incomes. In a generalized utilitarian social welfare function, each ϕ_i-function maps individual i's expected utility $U_i(\ell)$ to a utility of the impartial observer. These expected utilities differ across people, and concavity of the ϕ_i-functions is associated with egalitarianism over expected utilities, often called ex ante egalitarianism.[14]

We will show that concavity of the ϕ_i-functions is equivalent to an axiom that generalizes the example in the introduction. The example involved two indifference sets of the impartial observer, that containing (i, x_i) and (j, x_j) and that containing (i, x_j) and (j, x_i). We argued that a preference for fairness corresponds to preferring a randomization between these indifference sets in outcome lotteries to a randomization in identity lotteries. To generalize, suppose the impartial observer is indifferent between (z, ℓ') and (z', ℓ), and consider the product lottery (z, ℓ) that (in general) lies in a different indifference set. There are two ways to randomize between these indifference sets while remaining in the set of product lotteries. The product lottery $(z, \alpha\ell + (1-\alpha)\ell')$ randomizes between these indifference sets in outcome lotteries (i.e., real life chances); while the product lottery $(\alpha z + (1-\alpha)z', \ell)$ randomizes between these indifference sets in identity lotteries (i.e., imaginary accidents of birth).

Preference for Life Chances. *For any pair of identity lotteries z and z' in $\triangle(\mathcal{I})$, and any pair of outcome lotteries ℓ and ℓ' in $\triangle(\mathcal{X})$, if $(z, \ell') \sim (z', \ell)$ then $(z, \alpha\ell + (1-\alpha)\ell') \succsim (\alpha z + (1-\alpha)z', \ell)$ for all α in $(0, 1)$.*

If we add this axiom to the conditions of Theorem 1, then we obtain concave generalized utilitarianism.

[14] See for example, Broome (1984), Myerson (1981), Hammond (1981, 1982) and Meyer (1991). In our context, it is perhaps better to call this 'interim' egalitarianism since it refers to distributions 'after' the resolution of the identity lottery but 'before' the resolution of the outcome lottery. We can contrast this with a concern for *ex post* inequality of individuals' welfare, see for example Fleurbaey (2007).

Proposition 2 (Concavity) *Suppose that absence of unanimity applies. A generalized utilitarian impartial observer with representation* $\langle \{U_i, \phi_i\}_{i \in \mathcal{I}} \rangle$ *exhibits preference for life chances if and only if each of the* ϕ_i*-functions is concave.*

This result does rely on there being some richness in the underlying preferences so that preference for life chances has bite. In particular, example 2 in the supplementary appendix shows that, if all agents agree in their ranking of all outcome lotteries then the ϕ_i's need not be concave. This is ruled out in the proposition by absence of unanimity.

As discussed, Harsanyi treats identity and outcome lotteries as equivalent. Hence he implicitly imposes the following indifference.

Indifference between Life Chances and Accidents of Birth. *For any pair of identity lotteries* z *and* z' *in* $\triangle(\mathcal{I})$*, and any pair of outcome lotteries* ℓ *and* ℓ' *in* $\triangle(\mathcal{X})$*, if* $(z, \ell') \sim (z', \ell)$ *then* $(z, \alpha\ell + (1-\alpha)\ell') \sim (\alpha z + (1-\alpha)z', \ell)$ *for all* α *in* $(0, 1)$*.*

This is a very strong assumption. If we impose this indifference as an explicit axiom then, as a corollary of Proposition 2, we obtain that each ϕ_i-function must be affine. In this case, if we let $\hat{U}_i := \phi_i \circ U_i$, then $\hat{U}_i$ is itself a von Neumann-Morgenstern expected-utility representation of $\succsim_i$. Thus, we immediately obtain Harsanyi's utilitarian representation.

But, in fact, we obtain a stronger result. This indifference over the type of randomization allows us to dispense with the independence axiom over outcome lotteries for the individuals.

Theorem 3 (Utilitarianism) *Suppose that absence of unanimity applies. The impartial observer's preferences* $\succsim$ *admit a utilitarian representation* $\langle \{U_i\}_{i \in \mathcal{I}} \rangle$ *if and only if the impartial observer satisfies the acceptance principle, independence over identity lotteries, and is indifferent between life chances and accidents of birth.*

Moreover the functions U_i *are unique up to common positive affine transformations.*

Standard proofs of Harsanyi's utilitarianism directly impose stronger notions of independence.[15] For example:

[15] See section 6 for details.

Independence over Outcome Lotteries (for the Impartial Observer). *Suppose* (z,ℓ), $(z',\ell') \in \triangle(\mathcal{I}) \times \triangle(\mathcal{X})$ *are such that* $(z,\ell) \sim (z',\ell')$. *Then for all* $\tilde{\ell}$, $\tilde{\ell}' \in \triangle(\mathcal{X})$: $(z,\tilde{\ell}) \gtrsim (z',\tilde{\ell}')$ *if and only if* $(z, \alpha\tilde{\ell} + (1-a)\,\ell) \gtrsim (z', \alpha\tilde{\ell}' + (1-a)\,\ell')$ *for all* α *in* $(0,1]$.

This axiom is the symmetric analog of identity independence for the impartial observer reversing the roles of identity lotteries and outcome lotteries. Clearly, if the impartial observer satisfies this independence then it would be redundant for her to inherit independence over outcome lotteries from individual preferences; and moreover, given acceptance, this independence for the impartial observer imposes independence on the individuals. We do not directly impose independence over outcome lotteries on the impartial observer, but our axioms imply it.

Corollary 4 *Suppose that absence of unanimity applies. Then the impartial observer satisfies independence over outcome lotteries if she satisfies acceptance, independence over identity lotteries, and is indifferent between life chances and accidents of birth.*

To summarize: What separates Harsanyi from those generalized utilitarian impartial observers who are ex ante egalitarians are their preferences between outcome and identity lotteries. If the impartial observer prefers outcome lotteries, she is an ex ante egalitarian. If she is indifferent (like Harsanyi) then she is a utilitarian. Moreover, indifference between outcome and identity lotteries forces the generalized utilitarian to accept stronger notions of independence.

5 Different risk attitudes

Recall that an impartial observer's interpersonal welfare comparisons might rank $(i,x_i) \sim (j,x_j)$ and $(i,x_j) \sim (j,x_i)$, but if person i is more comfortable facing risk than person j, she might rank $\left(i, \frac{1}{2}[x_i] + \frac{1}{2}[x_j]\right) \succ \left(j, \frac{1}{2}[x_i] + \frac{1}{2}[x_j]\right)$. Harsanyi's utilitarianism rules this out.

An analogy might be useful. In the standard representative-agent model of consumption over time, each time period is assigned one utility function. This utility function must reflect both risk aversion in that period and substitutions between periods. Once utilities are scaled for inter-temporal welfare comparisons, there is limited scope to accommodate different risk attitudes across periods. Harsanyi's utilitarian impartial observer assigns one utility function per person.

This utility function must reflect both the risk aversion of that person and substitutions between people. Once utilities are scaled for interpersonal welfare comparisons, there is limited scope to accommodate different risk attitudes across people.

Given this analogy, it is not surprising that generalized utilitarianism can accommodate different risk attitudes. Each person is now assigned two functions, ϕ_i and u_i, so we can separate interpersonal welfare comparisons from risk aversion.

To be more precise, we first generalize the example in the introduction.

Similar Risks *Suppose the impartial observer assesses* $(i,\ell) \sim (j,\ell')$ *and* $(i,\tilde{\ell}) \sim (j,\tilde{\ell}')$. *Then, for all* α *in* $(0,1)$, *the two outcome lotteries* $\alpha\tilde{\ell} + (1-\alpha)\,\ell$ *and* $\alpha\tilde{\ell}' + (1-\alpha)\,\ell'$ *are* similar risks *for individuals* i *and* j *respectively.*

These risks are similar for i and j in that they are across outcome lotteries that the impartial observer has assessed to have equal welfare for individuals i and j respectively. If individual j is more risk averse than individual i, then we might expect the impartial observer to prefer to face these similar risks as person i.

Preference to Face Similar Risks as i **rather than** j *Fix a pair of individuals* i *and* j *in* $\mathcal{I}$. *The impartial observer is said to prefer to face similar risks as indivdiual* i *rather than as individual* j, *if any four outcome lotteries* ℓ,ℓ',$\tilde{\ell}$ *and* $\tilde{\ell}'$ *in* $\triangle(\mathcal{X})$, *such that* $(i,\ell) \sim (j,\ell')$ *and* $(i,\tilde{\ell}) \sim (j,\tilde{\ell}')$ *then,* $\left(i,\alpha\tilde{\ell} + (1-\alpha)\,\ell\right) \succsim \left(j,\alpha\tilde{\ell}' + (1-\alpha)\,\ell'\right)$ *for all* α *in* $[0,1]$.

Recall that agent j is more income risk averse than agent i if the function u_j that maps income to agent j's von Neumann-Morgenstern utility is a concave transformation of that function u_i for agent i; that is, $u_i \circ u_j^{-1}$ is convex. For each i, the function ϕ_i^{-1} maps the utilities of the impartial observer (used in her interpersonal welfare comparisons) to agent i's von Neumann-Morgenstern utility. Thus, if agent j is more (welfare) risk averse than agent i then ϕ_j^{-1} is a concave transformation of ϕ_i^{-1}; that is, $\phi_i^{-1} \circ \phi_j$ is convex everywhere where they are comparable. The next proposition makes this precise.

Proposition 5 (Different Risk Attitudes.) *Suppose that absence of unanimity applies. A generalized utilitarian impartial observer with representation* $\left\langle \{U_i, \phi_i\}_{i\in\mathcal{I}} \right\rangle$ *always prefers to face*

similar risks as i rather than j if and only if the composite function $\phi_i^{-1} \circ \phi_j$ is convex on the the domain $\mathcal{U}_{ji} := \{u \in \mathbb{R} : \text{there exists } \ell, \ell' \in \triangle(\mathcal{X}) \text{ with } (i,\ell) \sim (j,\ell') \text{ and } U_j(\ell') = u\}$.

Next consider indifference as to which individual should face similar risks.

Indifference between Individuals facing Similar Risks. *For any pair of individuals i and j in $\mathcal{I}$ and any four outcome lotteries ℓ,ℓ',$\tilde{\ell}$ and $\tilde{\ell}'$ in $\triangle(\mathcal{X})$, if $(i,\ell) \sim (j,\ell')$ and $(i,\tilde{\ell}) \sim (j,\tilde{\ell}')$ then, for all α in $[0,1]$, the impartial observer is indifferent between facing the similar risks $\alpha\tilde{\ell} + (1-\alpha)\ell$ and $\alpha\tilde{\ell}' + (1-\alpha)\ell'$ as individual i or j respectively.*

Harsanyi's utilitarian impartial observer satisfies this indifference: it is an immediate consequence of independence over outcome lotteries for the impartial observer. But we can imagine an impartial observer who, without necessarily satisfying all of Harsanyi's axioms, is nevertheless indifferent as to which individual should face similar risks. For example, consider an impartial observer in the analog of a 'representative-agent' model. In the standard representative-agent model, all individuals have the same preferences over private consumption and the same attitude to risk. In our setting, we must allow individuals to have different preferences over *public* outcomes.[16] But, as in the standard representative-agent model, we could assume that each individual had the same risk attitude across outcome lotteries that had been assessed to have equal welfare. This is precisely the indifference property above.

Given Proposition 5, for any two individuals i and j, indifference between individuals facing similar risks forces the ϕ_i and ϕ_j-functions to be identical up to positive affine transformations provided $\mathcal{U}_{ji}$ has a non-empty interior. Hence:

Proposition 6 (Common ϕ-Function) *Suppose that absence of unanimity applies and consider a generalized utilitarian impartial observer. There exists a generalized utilitarian representation $\langle \{U_i, \phi_i\}_{i \in \mathcal{I}} \rangle$ with $\phi_i = \phi$ for all i in $\mathcal{I}$ if and only if the impartial observer is indifferent between individuals facing similar risks.*

Moreover, if for any pair of individuals i and j in $\mathcal{I}$, there exists a sequence of individuals $j_1 \ldots j_N$ with $j_1 = i$ and $j_N = j$ such that $\mathcal{U}_{j_n j_{n-1}}$ has non-empty interior then the functions U_i

[16] For example, public outcome x_i might allocate an indivisible good to person i, while x_j might allocate it to person j.

are unique up to a common positive affine transformation, and the composite functions $\phi \circ U_i$ are unique up to a common positive affine transformation.

To compare results, a generalized utilitarian impartial observer who is not concerned about the issue of different individual risk attitudes (and hence satisfies indifference between individuals facing similar risks) need not be a utilitarian. She need only translate individuals' von Neumann-Morgenstern utilities using a common ϕ-function when making welfare comparisons across those individuals. Hence such an impartial observer can accommodate issues of fairness: in particular, the common ϕ-function might be concave.

In contrast, a generalized utilitarian impartial observer who is not concerned about issues of fairness (and hence satisfies indifference between life chances and accidents of birth) must be a utilitarian. Hence such an impartial observer cannot accommodate the issue of different individual risk attitudes.

To see this directly, recall that independence over outcome lotteries for the impartial observer immediately implies indifference between individuals facing similar risks. And, by corollary 4, for a generalized utilitarian impartial observer, indifference between life chances and accidents of birth implies independence over outcome lotteries for the impartial observer.

Consideration of different risk aversions and consideration of fairness are distinct issues and they may lead an impartial observer in opposite directions. For example, suppose that all individuals are extremely risk averse over outcome lotteries, but that the impartial observer is almost risk neutral over identity lotteries. This impartial observer, anticipating the real discomfort that outcome lotteries would cause people, might prefer to absorb the risk into the imaginary identity lottery of her thought experiment. That is, she might prefer a society in which most uncertainty has been resolved – and hence people would "know their fates" – by the time they were born. Such an impartial observer would prefer accidents of birth to life chances: she would be an ex ante *anti-egalitarian.*

6 Contrasting Independences and Domains

Recall that Harsanyi works with the full set of joint distributions $\triangle(\mathcal{I} \times \mathcal{X})$, not just the product lotteries $\triangle(\mathcal{I}) \times \triangle(\mathcal{X})$. He imposes independence directly on the impartial observer for all mixtures defined on that domain. In this section, we first consider the natural extensions of our axioms for the impartial observer in the larger domain $\triangle(\mathcal{I} \times \mathcal{X})$. Second, we consider restricting Harsanyi's original independence axiom defined on $\triangle(\mathcal{I} \times \mathcal{X})$ to the set of product lotteries $\triangle(\mathcal{I}) \times \triangle(\mathcal{X})$. Third, we discuss whether imposing identity and outcome independence directly on the impartial observer is enough to induce utilitarianism.

6.1 The full set of joint distributions.

Suppose that the impartial observer has preferences over the full space of joint distributions over identities and outcomes, $\triangle(\mathcal{I} \times \mathcal{X})$. With slight abuse of notation, let $\succsim$ continue to denote these larger preferences. For purposes of comparison, it is convenient to denote each element of $\triangle(\mathcal{I} \times \mathcal{X})$, in the form $\left(z, (\ell_i)_{i \in \mathcal{I}}\right)$ where $z \in \triangle(\mathcal{I})$ is the marginal on the identities and each $\ell_i \in \triangle(\mathcal{X})$ is the outcome lottery conditional on identity i obtaining. Thus $(\ell_i)_{i \in \mathcal{I}}$ is a vector of conditional outcome lotteries. Notice that, in this larger setting, the impartial observer imagines each individual having his own personal outcome lottery.

In this setting, the analog of our independence over identity lotteries axiom for the impartial observer is:

Constrained Independence over Identity Lotteries (for the Impartial Observer). *Suppose* $(z, (\ell_i)_{i \in \mathcal{I}})$, $(z', (\ell'_i)_{i \in \mathcal{I}}) \in \triangle(\mathcal{I} \times \mathcal{X})$ *are such that* $(z, (\ell_i)_{i \in \mathcal{I}}) \sim (z', (\ell'_i)_{i \in \mathcal{I}})$. *Then, for all* $\tilde{z}$, $\tilde{z}' \in \triangle(\mathcal{I})$: $(\tilde{z}, (\ell_i)_{i \in \mathcal{I}}) \succsim (\tilde{z}', (\ell'_i)_{i \in \mathcal{I}})$ *if and only if* $(\alpha\tilde{z} + (1-\alpha)\, z, (\ell_i)_{i \in \mathcal{I}}) \succsim (\alpha\tilde{z}' + (1-\alpha)\, z', (\ell'_i)_{i \in \mathcal{I}})$ *for all* α *in* $(0, 1]$.

This is the independence axiom suggested by Karni & Safra (2000).

Constrained independence over identity lotteries is weaker than Harsanyi's independence axiom in that it only applies to mixtures of identity lotteries. That is, like our independence axiom for the impartial observer, constrained independence over identity lotteries is independence for the

impartial observer over the lotteries that she faces directly – namely, identity lotteries – holding the vector of conditional outcome lotteries fixed. Notice, however, that each resolution of the identity lottery yields not just a different identity but also a different outcome lottery. This extends the bite of the axiom to the larger space $\triangle(\mathcal{I} \times \mathcal{X})$. When restricted to the set of product lotteries, $\triangle(\mathcal{I}) \times \triangle(\mathcal{X})$, conditional independence reduces to our independence axiom over identity lotteries.

The following axiom (also from Karni & Safra (2000)) is a slight strengthening of Harsanyi's acceptance axiom.

Acceptance* Principle. *For all i in $\mathcal{I}$, all $(\ell_1, \ldots, \ell_i, \ldots, \ell_I)$ in $\triangle(\mathcal{X})^I$ and ℓ_i' in $\triangle(\mathcal{X})$, $\ell_i \succsim_i \ell_i'$ if and only if $(i, (\ell_1, \ldots, \ell_i, \ldots, \ell_I)) \succsim (i, (\ell_1, \ldots, \ell_i', \ldots, \ell_I))$.*

The motivation for this axiom is the same as that for Harsanyi's axiom. The slight additional restriction is that, if the impartial observer knows that she will assume individual i's identity, she does not care about the (possibly different) conditional outcome lottery that she would have faced had she assumed some other identity.

If we replace our independence and acceptance axioms with these axioms, then our generalized utilitarian representation theorem holds exactly as stated in theorem 1 except that the representation becomes

$$V\left(z, (\ell_i)_{i \in \mathcal{I}}\right) = \sum_i z_i \phi_i \left(U_i\left(\ell_i\right)\right). \tag{3}$$

That is, each individual has a personal conditional outcome lottery ℓ_i in place of the common outcome lottery ℓ. The proof is essentially the same as that of theorem 1.[17] Moreover, proposition 2, theorem 3, proposition 5 and their corollaries all continue to hold (with the same modification about personal outcome lotteries) by the same proofs.[18] Thus, if we extend the analogs of our axioms to Harsanyi's setting $\triangle(\mathcal{I} \times \mathcal{X})$, we get essentially the same results.

[17] See the supplementary appendix. Alternatively, this generalized utilitarian representation could be obtained as a corollary of theorem 1 in Karni & Safra (2000).

[18] Corollary 4 also holds without this modification, and we can also obtain stronger versions of outcome independence.

6.2 Harsanyi's independence axiom restricted to product lotteries.

Conversely, now consider the restriction of Harsanyi's independence axiom to our setting, $\triangle(\mathcal{I}) \times \triangle(\mathcal{X})$. In this setting, the analog of Harsanyi's axiom is to apply independence to all mixtures that are well-defined in the set of product lotteries.[19] To understand how Harsanyi's independence relates to the axioms in this paper – and hence to see how Harsanyi implicitly imposes each of those axioms – it helps to unpack Harsanyi's independence axiom into three axioms each associated with the type of mixture to which it applies. First, Harsanyi's independence axiom restricted to product lotteries implies our independence over identity lotteries for the impartial observer. This independence axiom is also satisfied by our generalized utilitarian impartial observer. Second, it implies independence over outcome lotteries, imposed directly on the impartial observer not just derived via acceptance from the preferences of the individuals. This independence axiom immediately implies indifference between individuals facing similar risks.

Third, the restriction of Harsanyi's axiom also forces the impartial observer to apply independence to hybrid mixtures.

Independence over Hybrid Lotteries (for the Impartial Observer). *Suppose* $(z, \ell), (z', \ell') \in \triangle(\mathcal{I}) \times \triangle(\mathcal{X})$ *are such that* $(z, \ell) \sim (z', \ell')$. *Then for all* $\tilde{z} \in \triangle(\mathcal{I})$ *and all* $\tilde{\ell}' \in \triangle(\mathcal{X})$*:* $(\tilde{z}, \ell) \succsim$ (resp. $\precsim$) $(z', \tilde{\ell}')$ *if and only if* $(\alpha\tilde{z} + (1-\alpha)\, z, \ell) \succsim$ (resp. $\precsim$) $(z', \alpha\tilde{\ell}' + (1-a)\, \ell')$ *for all* α *in* $(0, 1]$.

In this axiom the lotteries being mixed on the left are identity lotteries (holding outcome lotteries fixed), while the lotteries being mixed on the right are outcome lotteries (holding identity lotteries fixed). This independence axiom immediately implies indifference between life chances and accidents of birth.

It follows from theorem 1 that, given absence of unanimity and acceptance, the first and third implication of Harsanyi's independence axiom when restricted to our setting, $\triangle(\mathcal{I}) \times \triangle(\mathcal{X})$ — i.e., identity and hybrid independence — are enough to yield Harsanyi's conclusion, utilitarianism.[20]

[19] This is the approach of Safra & Weissengrin (2002) who adapt Fishburn's (1982, chapter 7) work on product spaces of mixture sets.

[20] Given all three implications of Harsanyi's independence axiom (i.e., including outcome independence), we can dispense with absence of unanimity: see Safra & Weissengrin (2002).

6.3 Independence along both margins.

A natural question is whether we can replace hybrid independence with outcome independence in the statement above: that is, whether acceptance and both identity and outcome independence are enough to induce utilitarianism. We have argued in this paper that outcome independence is a strong assumption in the context of the impartial observer: it directly imposes independence over lotteries that she does not face directly, and by so doing implies much more than simply imposing independence on the individuals and acceptance on the impartial observer. Nevertheless, one might prefer such an axiomatization to using hybrid independence. First, hybrid independence might seem the least natural of the three implications of Harsanyi's independence axiom for product lotteries. Both outcome and identity independence only involve mixing one margin at a time. Second, an impartial observer might satisfy identity and outcome independence because she views the two types of randomization symmetrically – if independence applies to one margin then perhaps it should apply to the other – without taking a direct position on whether the two types of randomization are equivalent.

It turns out, however, that identity independence, outcome independence and acceptance are not enough to induce utilitarianism. In fact, we can see this using the example in the introduction. Once again, suppose that there are two individuals, i and j, and two states, x_i and x_j, denoting which agent is given a (possibly indivisible) good. As before, suppose that the impartial observer's preferences satisfy $(i, x_i) \sim (j, x_j)$ and $(i, x_j) \sim (j, x_i)$. Suppose that both individuals satisfy independence. Specifically, for any outcome lottery ℓ, player i's expected utility is given by $U_i(\ell) = \ell(x_i) - \ell(x_j)$ and player j's expected utility is given by $U_j(\ell) = \ell(x_j) - \ell(x_i)$. Let the impartial observer's preferences be given by the generalized utilitarian representation $V(z, \ell) := z_i \phi[U_i(\ell)] + z_j \phi[U_j(\ell)]$ where the (common) ϕ-function is given by:

$$\phi[u] = \begin{cases} u^k & \text{for } u \geq 0 \\ -(-u)^k & \text{for } u < 0 \end{cases}, \text{ for some } k > 0.$$

Since these preferences are generalized utilitarian, by theorem 1, they satisfy acceptance and identity independence. And since the ϕ-function is common, by proposition 6, they satisfy in-

difference between individuals facing similar risks. It is less obvious that they satisfy outcome independence but this is shown in the supplementary appendix.

These preferences even have the property (similar to utilitarianism) that if the impartial observer thinks she is equally likely to be either person, she is indifferent who gets the good. But these preferences do not satisfy utilitarianism unless $k = 1$. To see this, notice that these preferences fail indifference between life chances and accidents of birth. For example, we have $(i, x_i) \sim (j, x_j)$, but $(i, \alpha x_i + (1-\alpha) x_j) \nsim (\alpha [i] + (1-\alpha) [j], x_i)$ except in the special case when $\alpha = \frac{1}{2}$.

Nevertheless, the conjecture that independence along both margins implies utilitarianism is close to correct. Grant et al (2009: Theorem 7) show that, if there are three or more agents, under some richness conditions on the preferences, the combination of identity independence, outcome independence and acceptance do imply utilitarianism.

7 Knowledge, Agreement and Welfare

Two questions figure prominently in the debates on the impartial observer theorem. First, what is it that an individual imagines and knows when she imagines herself in the role of the impartial observer. Second, must all potential impartial obsevers agree in their preferences over extended lotteries. In this section, we consider four (of many) possible views on these questions and show how they relate to the issues of this paper: concern about different risk attitudes (loosely, does the impartial observer use a common ϕ-function); and concern about fairness (loosely, is her common ϕ-function affine).[21]

In one view of the impartial observer, she simply imagines being in the physical circumstances of person i or j facing the outcome lottery ℓ or ℓ'.[22] In this view, often associated with Vickrey (1945), the impartial observer does not attempt to imagine having person i's or j's preferences. In the context of our example, the impartial observer simply imagines herself having some chance of getting the indivisible good, and applies her own preferences about such outcome lotteries. Compared to other views, this approach does not require as much imagination or knowledge on

[21] The following builds especially on Weymark (1991) and Mongin (2001). For other views see for example d'Aspremont & Mongin (1998).

[22] Pattanaik (1968, p. 1155) and Harsanyi (1977, p. 52) refer to these as 'objective positions'.

behalf of the impartial observer. In particular, she need not know i's or j's preferences. If the impartial observer adopts this approach, loosely speaking, we get a common ϕ-function for free: the utilities in its domain are all utilities of the same agent, the impartial observer. The ϕ-function need not be affine however since the impartial observer might still, for example, prefer outcome to identity lotteries. In this approach, there is no reason to expect all impartial observers to agree. For example, different potential impartial observers will generally have different preferences over physical outcome lotteries. This approach does not attempt to follow the acceptance principle. Individuals' preferences over outcome lotteries (other than those of the impartial observer) play no role.

In a second view (the view taken in this paper), the impartial observer imagines not only being in the physical circumstances of person i or j but also adopting what Pattanaik (1968, p. 1155) calls "the subjective features of the respective individuals". Arrow (1963, p. 114, 1977) calls this "extended sympathy" but it is perhaps better to use Harsanyi's own term, "imaginative empathy":

> "This must obviously involve [her] imagining [her]self to be placed in individual i's *objective position*, i.e., to be placed in the objective positions (e.g., income, wealth, consumption level, state of health, social position) that i would face in social situation x. But it must also involve assessing these objective conditions in terms of i's own *subjective attitudes* and *personal preferences* ... rather than assessing them in terms of [her] own subjective attitudes and personal preferences. [Harsanyi, 1977, p. 52: notation changed to ours but emphasis in the original][23]

This approach requires more imagination and knowledge by the impartial observer; in particular, she is assumed to know the preferences of each individual over outcome lotteries and, by acceptance, to adopt these preferences when facing outcome lotteries as that individual. Knowledge and acceptance of individual preferences implies agreement across all potential impartial observers in

[23] Rawls also appeals to such imaginative empathy: "A competent judge ... must not consider his own *de facto* preferences as the necessarily valid measure of the actual worth of those interests which come before him, but ... be both able and anxious to determine, by imaginative appreciation, what those interests mean to persons who share them, and to consider them accordingly." (Rawls 1951, p. 179 quoted in Pattanaik 1968, p. 1157-8). See also Sen's (1979) behavioral and introspective bases for interpersonal comparisons of welfare.

ranking pairs of the form (i, ℓ) and (i, ℓ'). But, as Broome (1993) and Mongin (2001) have pointed out (and as Harsanyi (1977, p. 57) himself concedes), it does not imply agreement in ranking pairs of the form (i, ℓ) and (j, ℓ') where $i \neq j$. For example, each impartial observer can have her own rankings across others' subjective and objective positions.

Moreover, unlike in the Vickrey view above, a generalized utilitarian impartial observer in this setting need not use a common ϕ-function across all individuals. To see this, let us extend the example from the introduction by allowing the good being allocated to be divisible. Suppose that an impartial observer's own interpersonal assessments are such that she is indifferent between being person i with share s of the good and being person j with the same share s of the good. Suppose that for person i, the outome lottery $\frac{1}{2}x_i + \frac{1}{2}x_j$ in which he has a half chance of getting the whole good is indifferent to getting half the good for sure, but for person j this same lottery is indifferent to getting one third of the good for sure. Combining acceptance with her interpersonal assessments, the impartial observer must prefer facing this outcome lottery as person i. But, by proposition 6, this contradicts using a common ϕ-function (and in particular, not all the ϕ_i-functions can be affine).

A third, more welfarist view goes beyond the assumptions of this paper. Suppose that, when an impartial observer imagines being person i facing outcome lottery ℓ, she knows the (ex ante) 'welfare' that i attains from this lottery. That is, suppose that each person i has a commonly known 'welfare function' $w_i : \Delta(\mathcal{X}) \to \mathbb{R}$. If we assume what Weymark (1991) calls congruence between welfare and preference – that is, $\ell \succsim_i \ell'$ if and only if $w_i(\ell) \geq w_i(\ell')$ – then this implies, as before, that the impartial observer knows person i's preferences. But now suppose further that these welfares functions are at least ordinally measurable and fully comparable, and that the impartial observer satisfies the rule: $(i, \ell) \succsim (j, \ell')$ if and only if $w_i(\ell) \geq w_j(\ell')$. This extra assumption implies acceptance, but it is stronger. It implies that all potential impartial observers must agree in ranking pairs of the form (i, ℓ) and (j, ℓ').

Nevertheless, a generalized utilitarian impartial observer in this setting still need not use a common ϕ-function across all individuals. The example above still applies. The $w_i(\cdot)$ functions can encode the impartial observer's assessment about being indifferent between being i or j with

the same share s of the good; and they can encode i and j's different certainty equivalents. Again, this forces ϕ_i and ϕ_j to differ (and at least one to be non-affine).

Moreover, these welfarist assumptions still do not imply full agreement across potential impartial observers. All impartial observers must agree in the ranking of extended lotteries in which they know for sure which identity they will assume, but they can still differ in their ranking of general extended lotteries of the form (z, ℓ) and (z', ℓ'). For example, different impartial observers might have different preferences between outcome and identity lotteries. And/or each impartial observer can have her own risk attitude in facing identity lotteries, reflected in her own set of ϕ_i-functions. That is, even with these extreme assumptions, different impartial observers with different risk attitudes will make different social choices.

To get beyond this conclusion, a fourth view simply assumes that each potential impartial observer's von Neumann-Morgenstern utility $V(i, \ell)$ from the extended lottery (i, ℓ) is equal to the commonly known (fully comparable) welfare $w_i(\ell)$ which in turn is equal to individual i's von Neumann-Morgenstern utility $U_i(\ell)$.[24] In this case, all attitudes toward similar risks are the same; in particular, the preferences of the impartial observer and the individuals i and j in the example above can no longer apply. With this strong identification assumption, we finally get both an affine common ϕ-function (i.e., utilitarianism) and agreement among all potential impartial observers, but this approach seems a few assumptions beyond Harsanyi's claim to have derived utilitarianism from Bayesian rationality alone.

A Appendix: Proofs

We first establish some lemmas that will be useful in the proofs that follow.

The first lemma shows that, given absence of unanimity, we need at most two outcome lotteries, ℓ^1 and ℓ_2, to 'cover' the entire range of the impartial observer's preferences in the following sense: for all product lotteries (z, ℓ) either $(z, \ell) \sim \left(z', \ell^1\right)$ for some z', or $(z, \ell) \sim (z'', \ell_2)$ for some z'', or both. Moreover the set of product lotteries for which 'both' applies are not all indifferent.

To state this more formally, let the outcome lotteries ℓ^1, ℓ_2 (not necessarily distinct) and

[24] This identification is at the heart of the debate between Harsanyi and Sen. See Weymark (1991).

identity lotteries z^1, z_2 (not necessarily distinct) be such that $\left(z^1, \ell^1\right) \succ (z_2, \ell_2)$ and such that $\left(z^1, \ell^1\right) \succsim (z, \ell) \succsim (z_2, \ell_2)$ for all product lotteries (z, ℓ). That is, the product lottery $\left(z^1, \ell^1\right)$ is weakly better than all other product lotteries, and the product lottery (z_2, ℓ_2) is weakly worse than all other product lotteries. And let the identity lotteries z_1 and z^2 (not necessarily distinct) be such that $\left(z^1, \ell^1\right) \succsim \left(z, \ell^1\right) \succsim \left(z_1, \ell^1\right)$ for all product lotteries $\left(z, \ell^1\right)$, and $\left(z^2, \ell_2\right) \succsim (z, \ell_2) \succsim (z_2, \ell_2)$ for all product lotteries (z, ℓ_2). That is, given outcome lottery ℓ^1, the identity lottery z_1 is (weakly) worse than all other identity lotteries; and, given outcome lottery ℓ_2, the identity lottery z^2 is (weakly) better than all other identity lotteries. The existence of these special lotteries follows from continuity of $\succsim$, non-emptyness of $\succ$, and the compactness of $\Delta(\mathcal{I}) \times \Delta(\mathcal{X})$. Moreover, by independence over identity lotteries, we can take z^1, z_1, z^2, and z_2 each to be a degenerate identity lottery. Let these be i^1, i_1, i^2, and i_2 respectively.

Lemma 7 (Spanning) *Assume absence of unanimity applies and that the impartial observer satisfies acceptance and independence over identity lotteries. Let $i^1, i_1, i^2, i_2, \ell^1$, and ℓ_2 be defined as above. Then (a) either $\left(i_1, \ell^1\right) \sim (i_2, \ell_2)$, or $\left(i^2, \ell_2\right) \sim \left(i^1, \ell^1\right)$, or $\left(i^2, \ell_2\right) \succ \left(i_1, \ell^1\right)$. And (b), for any product lottery (z, ℓ), either $\left(i^1, \ell^1\right) \succsim (z, \ell) \succsim \left(i_1, \ell^1\right)$ or $\left(i^2, \ell_2\right) \succsim (z, \ell) \succsim (i_2, \ell_2)$ or both.*

Proof. (a) If $\ell^1 = \ell_2$, then the first two cases both hold. Otherwise, suppose that the first two cases do not hold; that is, $\left(i_1, \ell^1\right) \succ (i_2, \ell_2)$ and $\left(i^1, \ell^1\right) \succ \left(i^2, \ell_2\right)$. By the definition of i_1, we know that $\left(i_2, \ell^1\right) \succsim \left(i_1, \ell^1\right)$, and hence $\left(i_2, \ell^1\right) \succ (i_2, \ell_2)$. Using absence of unanimity and acceptance , there must exist another individual $\hat{\imath} \neq i_2$ such that $(\hat{\imath}, \ell_2) \succ \left(\hat{\imath}, \ell^1\right)$. Again by the definition of i_1, we know that $\left(\hat{\imath}, \ell^1\right) \succsim \left(i_1, \ell^1\right)$, and hence $(\hat{\imath}, \ell_2) \succ \left(i_1, \ell^1\right)$. By the definition of i^2, we know that $\left(i^2, \ell_2\right) \succsim (\hat{\imath}, \ell_2)$, and hence $\left(i^2, \ell_2\right) \succ \left(i_1, \ell^1\right)$, as desired. Part (b) follows immediately from (a). ■

The next lemma does not yet impose independence over outcome lotteries on individuals and hence yields a more general representation than that in theorem 1. The idea for this lemma comes from Karni & Safra (2000) but they work with the full set of joint distributions $\triangle(\mathcal{I} \times \mathcal{X})$ whereas we are restricted to the set of product lotteries $\triangle(\mathcal{I}) \times \triangle(\mathcal{X})$.

Lemma 8 (Affine Representation) *Suppose absence of unanimity applies. Then the impartial observer satisfies the acceptance principle and independence over identity lotteries if and only if there exist a continuous function* $V : \triangle(\mathcal{I}) \times \triangle(\mathcal{X}) \to \mathbb{R}$ *that represents* $\succsim$*, and, for each individual* i *in* $\mathcal{I}$*, a function* $V_i : \triangle(\mathcal{X}) \to \mathbb{R}$*, that represents* $\succsim_i$*, such that for all* (z, ℓ) *in* $\triangle(\mathcal{I}) \times \triangle(\mathcal{X})$*,*

$$V(z, \ell) = \sum_{i=1}^{I} z_i V_i(\ell). \tag{4}$$

Moreover the functions V_i *are unique up to common positive affine transformations.*

Proof . Since the representation is affine in identity lotteries, it is immediate that the represented preferences satisfy the axioms. We will show that the axioms imply the representation.

Let $i^1, i_1, i^2, i_2, \ell^1$, and ℓ_2 be defined as in lemma 7 above. Given continuity, an immediate consequence of lemma 7 is that, for any product lottery (z, ℓ), either $(z, \ell) \sim \left(z', \ell^1\right)$ for some z', or $(z, \ell) \sim (z'', \ell_2)$ for some z'' or both. Moreover, we can choose the z' such that its support only contains individuals i^1 and i_1. And similarly for z'' with respect to i^2 and i_2.

The proof of lemma now proceeds with two cases.

Case (1) The easiest case to consider is where $\ell^1 = \ell_2$. In this case, $\left(i^1, \ell^1\right) \succ \left(i_1, \ell^1\right)$, and $\left(i^1, \ell^1\right) \succsim (z, \ell) \succsim \left(i_1, \ell^1\right)$, for all (z, ℓ). Then, for each (z, ℓ), let $V(z, \ell)$ be defined by

$$\left(V(z, \ell)\left[i^1\right] + (1 - V(z, \ell))\left[i_1\right], \ell^1\right) \sim (z, \ell).$$

By continuity and independence over identity lotteries, such a $V(z, \ell)$ exists and is unique.

To show that this representation is affine, notice that if $\left(V(z, \ell)\left[i^1\right] + (1 - V(z, \ell))\left[i_1\right], \ell^1\right) \sim (z, \ell)$ and $\left(V(z', \ell)\left[i^1\right] + (1 - V(z', \ell))\left[i_1\right], \ell^1\right) \sim (z', \ell)$ then independence over identity lotteries implies $\left(\left[\alpha V(z, \ell) + (1 - \alpha) V(z', \ell)\right]\left[i^1\right] + \left[1 - \alpha V(z, \ell) - (1 - \alpha) V(z', \ell)\right]\left[i_1\right], \ell^1\right) \sim (\alpha z + (1 - \alpha) z', \ell)$. Hence $\alpha V(z, \ell) + (1 - \alpha) V(z', \ell) = V(\alpha z + (1 - \alpha) z', \ell)$.

Since any identity lottery z in $\Delta(\mathcal{I})$ can be written as $z = \sum_i z_i [i]$, proceeding sequentially on $\mathcal{I}$, affinity implies $V(z, \ell) = \sum_i z_i V(i, \ell)$. Finally, by acceptance, $V(i, \cdot)$ agrees with $\succsim_i$ on $\Delta(\mathcal{X})$. Hence, if we define $V_i : \Delta(\mathcal{X}) \to \mathbb{R}$ by $V_i(\ell) = V(i, \ell)$, then V_i represents individual i's preferences. The uniqueness argument is standard: see for example, Karni & Safra (2000, p. 321).

Case (2). If $\left(i_1, \ell^1\right) \sim (i_2, \ell_2)$ then $\left(i^1, \ell^1\right) \succsim (z, \ell) \succsim \left(i_1, \ell^1\right)$ for all (z, ℓ) and hence case (1) applies. Similarly, if $\left(i^2, \ell_2\right) \sim \left(i^1, \ell^1\right)$ then $\left(i^2, \ell_2\right) \succsim (z, \ell) \succsim (i_2, \ell_2)$ for all (z, ℓ), and again case (1) applies (with ℓ_2 in place of ℓ^1). Hence suppose that $\left(i^1, \ell^1\right) \succ \left(i^2, \ell_2\right)$ and that $\left(i_1, \ell^1\right) \succ (i_2, \ell_2)$. Then, by lemma 7, $\left(i^1, \ell^1\right) \succ \left(i^2, \ell_2\right) \succ \left(i_1, \ell^1\right) \succ (i_2, \ell_2)$; that is, we have two overlapping intervals that 'span' the entire range of the impartial observer's preferences.

Then, just as in case (1), we can construct an affine function $V^1(\cdot, \cdot)$ to represent the impartial observer's preferences $\succsim$ restricted to those (z, ℓ) such that $\left(i^1, \ell^1\right) \succsim (z, \ell) \succsim \left(i_1, \ell^1\right)$, and we can construct an affine function $V^2(\cdot, \cdot)$ to represent $\succsim$ restricted to those (z, ℓ) such that $\left(i^2, \ell_2\right) \succsim (z, \ell) \succsim (i_2, \ell_2)$. We can then apply an affine re-normalization of either V_1 or V_2 such the (re-normalized) representations agree on the 'overlap' $\left(i^2, \ell_2\right) \succsim (z, \ell) \succsim \left(i_1, \ell^1\right)$. Since $V_1(\cdot, \cdot)$ and $V_2(\cdot, \cdot)$ are affine, the re-normalized representation is affine, and induction on I (plus acceptance) gives us $V(z, \ell) = \sum_i z_i V_i(\ell)$ as before. Again, uniqueness follows from standard arguments. ■

Remark. The argument in case (1) above is similar to that in Safra & Weisengrin (2003, p. 184) and Karni & Safra (2000, p. 320) except that, in the latter case, the analog of ℓ^1 is a vector of outcome lotteries, with a different outcome lottery for each agent. Both these papers use stronger axioms to obtain a unique representation when case (1) does not apply. Our argument for these cases applies lemma 7 which in turn uses the richness condition, absence of unanimity, in place of any stronger axiom on the preferences of the impartial observer.

Proof of Theorem 1 (Generalized Utilitarianism): It is immediate that the represented preferences satisfy the axioms. We will show that the axioms imply the representation. If we add to lemma 8 (the affine representation lemma) the assumption that each individual satisfies independence over outcome lotteries, then it follows immediately that each V_i-function in representation (4) must be a strictly increasing transformation, ϕ_i, of a von Neumann-Morgenstern expected-utility representation, U_i. Thus, we obtain a generalized utilitarian representation. ■

Proof of Proposition 2 (Concavity) For each i in $\mathcal{I}$, set $V_i(\ell) := V(i, \ell) = \phi_i[U_i(\ell)]$ for all ℓ. That is, these are the V_i's from the affine representation in lemma 8. Since each U_i is affine in outcome lotteries, each $V(i, \cdot)$ is concave in outcome lotteries if and only if the corresponding ϕ_i

is concave.

To show that concavity is sufficient, suppose $(z, \ell') \sim (z', \ell)$. Using the representation in lemma 8 and imposing concavity, we obtain $V(z, \alpha\ell + (1-\alpha)\ell') = \sum_{i=1}^{I} z_i V_i(\alpha\ell + (1-\alpha)\ell') = \sum_{i=1}^{I} z_i V(i, \alpha\ell + (1-\alpha)\ell') \geq \sum_{i=1}^{I} z_i[\alpha V(i,\ell) + (1-\alpha)V(i,\ell')] = \alpha V(z,\ell) + (1-\alpha)V(z,\ell')$. Using the fact that $(z, \ell') \sim (z', \ell)$, the last expression is equal to $\alpha V(z,\ell) + (1-\alpha)V(z',\ell) = V(\alpha z + (1-\alpha)z', \ell)$. Hence the impartial observer exhibits a preference for life chances.

For necessity, we need to show that for all i and all ℓ, $\ell' \in \triangle(\mathcal{X})$, $V(i, \alpha\ell + (1-\alpha)\ell') \geq \alpha V(i,\ell) + (1-\alpha)V(i,\ell')$ for all α in $[0,1]$. So let $\succsim$ exhibit preference for life chances, fix i and consider ℓ, $\ell' \in \triangle(\mathcal{X})$. Assume first that $\ell \sim_i \ell'$. By acceptance, $V(i,\ell) = V(i,\ell')$. Hence, by preference for life chances,

$$\begin{aligned}
& V(i, \alpha\ell + (1-\alpha)\ell') \\
\geq\ & V(\alpha[i] + (1-\alpha)[i], \ell) \quad \text{(by preference for life chances)} \\
=\ & V(i,\ell) \\
=\ & \alpha V(i,\ell) + (1-\alpha)V(i,\ell') \quad \text{(since } V(i,\ell) = V(i,\ell')\text{)},
\end{aligned}$$

as desired.

Assume henceforth that $\ell \succ_i \ell'$ (and, by acceptance, $V(i,\ell) > V(i,\ell')$). By absence of unanimity, there must exist a j such that $V(j,\ell) < V(j,\ell')$. There are three cases to consider.

(a) If $V(i,\ell') \geq V(j,\ell)$ then, by the representation in lemma 8, there exists z' (of the form $\beta[i] + (1-\beta)[j]$) such that $V(z',\ell) = V(i,\ell')$. Thus, for all α in $(0,1)$,

$$\begin{aligned}
& V(i, \alpha\ell + (1-\alpha)\ell') \\
\geq\ & V(\alpha[i] + (1-\alpha)z', \ell) \quad \text{(by preference for life chances)} \\
=\ & \alpha V(i,\ell) + (1-\alpha)V(z',\ell) \\
=\ & \alpha V(i,\ell) + (1-\alpha)V(i,\ell') \quad \text{(since } V(z',\ell) = V(i,\ell')\text{)},
\end{aligned}$$

as desired.

Assume henceforth that $V(j,\ell) > V(i,\ell')$ (which implies $V(j,\ell') > V(i,\ell')$).

(b) If $V(j,\ell') \geq V(i,\ell)$ then, by the representation in lemma 8, there exists z (of the form $\beta[i] + (1-\beta)[j]$) such that $V(z,\ell') = V(i,\ell)$. Thus, for all α in $(0,1)$,

$$\begin{aligned} & V(i, \alpha\ell' + (1-\alpha)\ell) \\ \geq\ & V(\alpha[i] + (1-\alpha)z, \ell') \quad \text{(by preference for life chances)} \\ =\ & \alpha V(i,\ell') + (1-\alpha)V(z,\ell') \\ =\ & \alpha V(i,\ell') + (1-\alpha)V(i,\ell) \quad (\text{since } V(z,\ell') = V(i,\ell)), \end{aligned}$$

as desired.

(c) Finally, let $V(i,\ell) > V(j,\ell') > V(j,\ell) > V(i,\ell')$. By the continuity of V, there exist β^0, β_0 in $(0,1)$ such that $\beta^0 > \beta_0$, and such that $V(i, \beta^0\ell + (1-\beta^0)\ell') = V(j,\ell')$ and $V(i, \beta_0\ell + (1-\beta_0)\ell') = V(j,\ell)$. Denote $\ell_0 = \beta_0\ell + (1-\beta_0)\ell'$. Then, similarly to part (a),

$$V_i(\gamma\ell + (1-\gamma)\ell_0) \geq \gamma V_i(\ell) + (1-\gamma)V_i(\ell_0)$$

for all $\gamma \in (0,1)$. Next, denote $\ell^0 = \beta^0\ell + (1-\beta^0)\ell'$. Then, similarly to part (b),

$$V_i(\gamma\ell' + (1-\gamma)\ell^0) \geq \gamma V_i(\ell') + (1-\gamma)V_i(\ell^0)$$

for all $\gamma \in (0,1)$. Therefore, restricted to the line segment $[\ell', \ell]$, the graph of V_i lies weakly above the line connecting $(\ell', V_i(\ell'))$ and $(\ell^0, V_i(\ell^0))$ (as does the point $(\ell_0, V_i(\ell_0))$) and weakly above the line connecting $(\ell_0, V_i(\ell_0))$ and $(\ell, V_i(\ell))$ (as does the point $(\ell^0, V_i(\ell^0))$). Hence, $V_i(\alpha\ell + (1-\alpha)\ell') \geq \alpha V_i(\ell) + (1-\alpha)V_i(\ell')$ for all $\alpha \in (0,1)$. ■

Proof of Theorem 3 (Utilitarianism): It is immediate that the represented preferences satisfy the axioms. We will show that the axioms imply the representation. Given acceptance, the proof of proposition 2 (concavity) shows that the impartial observer satisfies preference for life chances if and only if, each V_i in the representation in lemma 8 is concave in outcome lotteries. Notice, in particular, that this argument never uses the fact that each individual satisfies independence over outcome lotteries. By a similar argument, the impartial observer is indifferent between life chances and accidents of birth if and only if each V_i is affine in outcome lotteries. To complete

the representation, for each i, set $U_i(\cdot) \equiv V_i(\cdot)$ to obtain the required von Neumann-Morgenstern expected-utility representation of individual i's preferences $\succsim_i$. ■

Proof of Corollary 4 (Outcome Independence). This result can be obtained as a corollary of theorem 3 (Utilitarianism I). Alternatively, the proof of proposition 2 (concavity) shows that the impartial observer is indifferent between life chances and accidents of birth if and only if, for all i in $\mathcal{I}$, $V(i,\cdot)$ is affine in outcome lotteries. Using the representation in lemma 8, we obtain $V(z, \alpha\ell + (1-\alpha)\ell') = \sum_{i=1}^{I} z_i V(i, \alpha\ell + (1-\alpha)\ell') = \sum_{i=1}^{I} z_i[\alpha V(i,\ell) + (1-\alpha) V(i,\ell')] = \alpha V(z,\ell) + (1-\alpha) V(z,\ell')$. That is, the impartial observer is affine in outcome lotteries. Hence it follows that the impartial observer satisfies independence over outcome lotteries. ■

Proof of Proposition 5 (Different Risk Attitudes). First, notice that if $\mathcal{U}_{ji}$ is not empty then it is a closed interval. If $\mathcal{U}_{ji}$ has an empty interior then the proposition holds trivially true. Therefore, assume that $\mathcal{U}_{ji} = [\underline{u}_{ji}, \bar{u}_{ji}]$ where $\underline{u}_{ji} < \bar{u}_{ji}$.

To prove that $\phi_i^{-1} \circ \phi_j$ convex is sufficient, fix $\ell, \ell', \tilde{\ell}$ and $\tilde{\ell}'$ such that $V(i,\ell) = V(j,\ell')$ and $V\left(i,\tilde{\ell}\right) = V\left(j,\tilde{\ell}'\right)$. We want to show that $V\left(i, \alpha\tilde{\ell} + (1-\alpha)\ell\right) \geq V\left(j, \alpha\tilde{\ell}' + (1-\alpha)\ell'\right)$. By construction, both $U_j(\ell')$ and $U_j\left(\tilde{\ell}'\right)$ lie in $\mathcal{U}_{ji}$. Moreover, we have $U_i(\ell) = \phi_i^{-1} \circ \phi_j [U_j(\ell')]$ and $U_i\left(\tilde{\ell}\right) = \phi_i^{-1} \circ \phi_j \left[U_j\left(\tilde{\ell}'\right)\right]$. Applying the representation we obtain,

$$
\begin{aligned}
& V\left(i, \alpha\tilde{\ell} + (1-\alpha)\ell\right) \\
&= \phi_i\left[U_i\left(\alpha\tilde{\ell} + (1-\alpha)\ell\right)\right] && \text{(by the representation)} \\
&= \phi_i\left[\alpha U_i\left(\tilde{\ell}\right) + (1-\alpha) U_i(\ell)\right] && \text{(by affinity of } U_i\text{)} \\
&= \phi_i\left[\alpha\phi_i^{-1} \circ \phi_j\left[U_j\left(\tilde{\ell}'\right)\right] + (1-\alpha)\phi_i^{-1} \circ \phi_j [U_j(\ell')]\right] && \text{(by the representation)} \\
&\geq \phi_i\left[\phi_i^{-1} \circ \phi_j\left[\alpha U_j\left(\tilde{\ell}'\right) + (1-\alpha) U_j(\ell')\right]\right] && \text{(by convexity of } \phi_i^{-1} \circ \phi_j\text{)} \\
&= \phi_j\left[U_j\left(\alpha\tilde{\ell}' + (1-\alpha)\ell'\right)\right] && \text{(by affinity of } U_j\text{)} \\
&= V\left(j, \alpha\tilde{\ell}' + (1-\alpha)\ell'\right) && \text{(by the representation)}
\end{aligned}
$$

To prove that $\phi_i^{-1} \circ \phi_j$ convex is necessary, fix v,w in $\mathcal{U}_{ji}$. By the definition of $\mathcal{U}_{ji}$, there exist outcome lotteries $\ell, \ell' \in \triangle(\mathcal{X})$ such that $U_j(\ell') = v$ and $U_i(\ell) = \phi_i^{-1} \circ \phi_j(v)$; and there exist

outcome lotteries $\tilde{\ell}, \tilde{\ell}' \in \triangle(\mathcal{X})$ such that $U_j\left(\tilde{\ell}'\right) = w$ and $U_i\left(\tilde{\ell}\right) = \phi_i^{-1} \circ \phi_j(w)$. By construction, we have $V(i, \ell) = V(j, \ell')$ and $V\left(i, \tilde{\ell}\right) = V\left(j, \tilde{\ell}'\right)$. Therefore, for all α in $(0, 1)$

$$\begin{aligned}
\phi_i\left[U_i\left(\alpha\tilde{\ell} + (1-\alpha)\ell\right)\right] &\geq \phi_j\left[U_j\left(\alpha\tilde{\ell}' + (1-\alpha)\ell'\right)\right] \quad \Rightarrow \\
\alpha U_i\left(\tilde{\ell}\right) + (1-\alpha) U_i(\ell) &\geq \phi_i^{-1} \circ \phi_j\left[\alpha U_j\left(\tilde{\ell}'\right) + (1-\alpha) U_j(\ell')\right] \quad \Rightarrow \\
\alpha\phi_i^{-1} \circ \phi_j(w) + (1-\alpha)\phi_i^{-1} \circ \phi_j(v) &\geq \phi_i^{-1} \circ \phi_j(\alpha w + (1-\alpha) v)
\end{aligned}$$

Since v and w were arbitrarily, the last inequality corresponds to the convexity of $\phi_i^{-1} \circ \phi_j$ on $\mathcal{U}_{ji}$.

■

Proof of Proposition 6 (Common ϕ-function) . Necessity follows immediately from proposition 5. For the sufficiency argument, first fix a representation $\left\langle \{U_i, \phi_i\}_{i \in \mathcal{I}} \right\rangle$ of the preferences of the generalized utilitarian impartial observer. Recall that, by theorem 1, the composite functions $\phi_i \circ U_i$ are unique up to a common positive affine transformation. The argument proceeds by a series of steps to construct a new representation $\left\langle \left\{\hat{U}_i, \hat{\phi}_i\right\}_{i \in \mathcal{I}} \right\rangle$ with $\hat{\phi}_i \equiv \phi$ for all i in $\mathcal{I}$. The construction leaves the composite functions unchanged; that is, $\phi_i \circ U_i \equiv \phi \circ \hat{U}_i$ for all i. To start, let the outcome lottery ℓ^1 and the individual i^1 be such that $\left(i^1, \ell^1\right) \succsim (j, \ell')$ for all individuals $j \in \mathcal{I}$ and outcome lotteries ℓ' in $\triangle(\mathcal{X})$.

Step 1. Suppose there exists a second individual j such that the interval $\mathcal{U}_{ji^1}$ has a non-empty interior. By Proposition 5, if the impartial observer is indifferent between facing similar risks as i^1 or j, then $\phi_{i^1}^{-1} \circ \phi_j$ is affine on $\mathcal{U}_{ji^1}$. Since $\mathcal{U}_{ji^1}$ has a non-empty interior, $\phi_{i^1}^{-1} \circ \phi_j$ has a unique affine extension on $\mathbb{R}$. Define a new von Neumann-Morgenstern utility function $\hat{U}_j$ for agent j by the affine transformation, $\hat{U}_j(\ell) := \phi_{i^1}^{-1} \circ \phi_j[U_j(\ell)]$ for all ℓ in $\triangle(\mathcal{X})$. Define a new transformation function $\hat{\phi}_j$ for agent j by setting $\hat{\phi}_j\left(\hat{U}_j(\ell)\right) := \phi_j(U_j(\ell))$. Thus, in particular, if $(i^1, \ell) \sim (j, \ell')$ (and hence $\phi_j[U_j(\ell')] = \phi_{i^1}[U_{i^1}(\ell)]$), then by construction we have $\hat{U}_j(\ell') = U_{i^1}(\ell)$. Moreover, by construction, we have $\hat{\phi}_j(u) = \phi_{i^1}(u)$ for all u in the intersection of the ranges $U_i(\triangle(\mathcal{X})) \cap \hat{U}_j(\triangle(\mathcal{X}))$. Hence, with slight abuse of notation we can write $\phi := \hat{\phi}_j = \phi_{i^1}$, even if this extends the domain of ϕ_{i^1}. Thus, we can construct a new generalized utilitarian representation of the same preferences with U_j replaced by $\hat{U}_j$ and ϕ_j replaced by ϕ in which the two individuals i^1 and

j share a common ϕ. Uniqueness of the U_i up to common positive affine transformations holds because, by construction, $(i^1, \ell) \sim (j, \ell')$ implies $U_{i^1}(\ell) = \hat{U}_j(\ell')$.

Step 2. By repeating step 1, for any individual j' in $\mathcal{I}$ such that there exists a sequence of individuals $j_1 \ldots j_N$ with $j_1 = i^1$ and $j_N = j'$ such that $\mathcal{U}_{j_n j_{n-1}}$ has non-empty interior, we can construct a new generalized utilitarian representation in which the two individuals i^1 and j' share a common ϕ. Let $\mathcal{I}^1$ be the set of individuals who can be connected to i^1 in this manner. If $\mathcal{I}^1 = \mathcal{I}$, then we are done.

Step 3. Suppose then that $\mathcal{I} \backslash \mathcal{I}^1$ is non-empty. By construction, $(j, \ell'') \gtrsim (j', \ell')$ for all ℓ', ℓ'' in $\triangle(\mathcal{X})$ and all $j \in \mathcal{I}_1$ and $j' \in \mathcal{I} \backslash \mathcal{I}_1$. Let $i' \in \mathcal{I} \backslash \mathcal{I}^1$ and $\hat{\ell} \in \triangle(\mathcal{X})$ be such that $\left(i', \hat{\ell}\right) \gtrsim (j', \ell')$ for all individuals $j' \in \mathcal{I} \backslash \mathcal{I}^1$ and outcome lotteries ℓ' in $\triangle(\mathcal{X})$. If $(j, \ell'') \sim (i', \ell)$ for some ℓ, ℓ'' in $\triangle(\mathcal{X})$ and $j \in \mathcal{I}_1$: let $\hat{U}_{i'}$ be a positive affine transformation of $U_{i'}$ such that $\hat{U}_{i'}(\ell) = U_j(\ell'')$, and let $\hat{\phi}_{i'}$ be such that $\hat{\phi}_{i'} \circ \hat{U}_{i'} \equiv \phi_{i'} \circ U_{i'}$. Then simply extend ϕ on the range of $\hat{U}_{i'}$ by setting $\phi := \hat{\phi}_{i'}$. Conversely, if $(j, \ell'') \succ (i', \ell)$ for all ℓ, ℓ'' in $\triangle(\mathcal{X})$ and $j \in \mathcal{I}_1$: let $\hat{U}_{i'}$ be a positive affine transformation of $U_{i'}$ such that $\hat{U}_{i'}(\ell) < U_j(\ell'')$ for all ℓ, ℓ'' in $\triangle(\mathcal{X})$ and $j \in \mathcal{I}_1$, and let $\hat{\phi}_{i'}$ be such that $\hat{\phi}_{i'} \circ \hat{U}_{i'} \equiv \phi_{i'} \circ U_{i'}$. Again, extend ϕ on the range of $\hat{U}_{i'}$ by setting $\phi := \hat{\phi}_{i'}$.

Step 4. Repeat steps 1 and 2 using i' in place of i^1 and ϕ in place of ϕ_{i^1}. Let $\mathcal{I}^2$ be the set of individuals who can be connected to i' when step 2 is repeated. Notice that, by construction $\mathcal{I}^1 \cap \mathcal{I}^2$ is empty. If $\mathcal{I}^1 \cup \mathcal{I}^2 = \mathcal{I}$ then we are done. If $\mathcal{I}^1 \cup \mathcal{I}^2 \neq \mathcal{I}$ then repeat step 3. Let i'' be the individual in $I \backslash \left(\mathcal{I}^1 \cup \mathcal{I}^2\right)$ that corresponds to i' in this step. Then repeat steps 1 and 2 using i'' place of i. From the finiteness of $\mathcal{I}$, this process can be repeated only a finite number of times before we exhaust $\mathcal{I}$. ■

B Supplementary Appendix:

This appendix contains two counter-examples mentioned in the text and also the key step to show that the proof of theorem 1 extends to obtain the form of generalized utilitarian representation given in expression (3) for preferences defined on $\triangle(\mathcal{I} \times \mathcal{X})$ and the corresponding axioms as given in section 6.

Examples. For each of the following examples, let $\mathcal{I} = \{1, 2\}$ and $\mathcal{X} = \{x_1, x_2\}$. To simplify

notation, for each $z \in \triangle(\mathcal{I})$, let $q = z_2$; and for each $\ell \in \triangle(\mathcal{X})$ let $p := \ell(x_2)$. Then, with slight abuse of notation, we write $(q,p) \succsim (q',p')$ for $(z,\ell) \succsim (z',\ell')$, and write $V(q,p)$ for $V(z,\ell)$.

Example 1 simply translates the example discussed in section 6 to show that the impartial observer might satisfy acceptance, and both identity and outcome independence but not be utilitarian.

Example 1 *Let agent 1's preferences be given by* $U_1(p) = (1-2p)$*, and let agent 2's preferences be given by* $U_2(p) = (2p-1)$*. Let the impartial observer's preferences be given by* $V(q,p) := (1-q)\,\phi[U_1(p)] + q\phi[U_2(p)]$*, where the (common)* ϕ*-function is given by:*

$$\phi[u] = \begin{cases} u^k & \text{for } u \geq 0 \\ -(-u)^k & \text{for } u < 0 \end{cases} \text{, for some } k > 0$$

Acceptance and identity independence were discussed in the text. To show that this example satisfies outcome independence, consider the inverse function $\phi^{-1}(u) = u^{1/k}$ for $u \geq 0$ and $\phi^{-1}(u) = -(-u)^{1/k}$ for $u < 0$. This is a strictly increasing function. Therefore, the function $\phi^{-1}[V(\cdot,\cdot)]$ represents the same preferences as $V(\cdot,\cdot)$.

It is enough to show that we can write

$$\phi^{-1}[V(q,p)] = (1-p)\,\phi^{-1}[(1-2q)] + p\phi^{-1}[(2q-1)].$$

This alternative representation is symmetric to the original representation $V(\cdot,\cdot)$ with the p's and q's reversed and ϕ^{-1} replacing ϕ. Since the alternative representation is affine in p, preferences must satisfy independence over outcome lotteries.

To confirm that $\phi^{-1}[V(\cdot,\cdot)]$ takes this form, it is instructive to rewrite $V(q,p)$ as follows:

$$\begin{aligned} V(q,p) &= \begin{cases} (1-2q)(1-2p)^k & \text{for } p < 1/2 \\ (2q-1)(2p-1)^k & \text{for } p > 1/2 \end{cases} \\ &= \begin{cases} (1-2q)(1-2p)^k & \text{for } q < 1/2,\ p < 1/2 \text{ (and } V(q,p) > 0) \\ -(2q-1)(1-2p)^k & \text{for } q > 1/2,\ p < 1/2 \text{ (and } V(q,p) < 0) \\ 0 & \text{for } (2q-1)(2p-1) = 0 \\ -(1-2q)(2p-1)^k & \text{for } q < 1/2,\ p > 1/2 \text{ (and } V(q,p) < 0) \\ (2q-1)(2p-1)^k & \text{for } q > 1/2,\ p > 1/2 \text{ (and } V(q,p) > 0) \end{cases} . \end{aligned}$$

Hence,

$$\begin{aligned}
\phi^{-1}\circ V(q,p) &= \begin{cases}
(1-2q)^{1/k}(1-2p) & \text{for } q<1/2,\ p<1/2\\
-(2q-1)^{1/k}(1-2p) & \text{for } q>1/2,\ p<1/2\\
0 & \text{for } (2q-1)(2p-1)=0\\
-(1-2q)^{1/k}(2p-1) & \text{for } q<1/2,\ p>1/2\\
(2q-1)^{1/k}(2p-1) & \text{for } q>1/2,\ p>1/2
\end{cases}\\
&= \begin{cases}
(1-p)\left[(1-2q)^{1/k}\right]+p\left[-\left[-(2q-1)\right]^{1/k}\right] & \text{for } q<1/2\\
0 & \text{for } q=1/2\\
(1-p)\left[-\left[-(1-2q)\right]^{1/k}\right]+p\left[(2q-1)^{1/k}\right] & \text{for } q>1/2
\end{cases}\\
&= (1-p)\,\phi^{-1}\left[(1-2q)\right]+p\phi^{-1}\left[(2q-1)\right]
\end{aligned}$$

which equals $(1-p)\,\phi^{-1}\left[(1-2q)\right]+p\phi^{-1}\left[(2q-1)\right]$ as desired. ■

Example 2 shows that the impartial observer's preferences can satisfy all the conditions of proposition 2 (the concavity result) except absence of unanimity and yet the functions ϕ_i need not be concave. That is, absence of unanimity is essential.

Example 2 *Let the individual's preferences be given by* $U_1(p)=U_2(p)=p$, *and let the impartial observer's preferences be given by* $V(q,p):=(1-q)\,\phi_1\left[U_1(p)\right]+q\phi_2\left[U_2(p)\right]$ *where*

$$\phi_1(u):=\begin{cases}1/4+u/2 & \textit{for } u\le 1/2\\ u & \textit{for } u>1/2\end{cases}$$
$$\phi_2(u):=\begin{cases}u & \textit{for } u\le 1/2\\ 2u-1/2 & \textit{for } u>1/2\end{cases}$$

Since $U_1=U_2$, both individuals have the same ranking over outcome lotteries and so the impartial observer's preferences violate absence of unanimity. Clearly, the functions $\phi_1(.)$ and $\phi_2(.)$ are not concave. To see that the impartial observer satisfies preference for life chances, without loss of generality let $p\le p'$ and notice that $(q,p')\sim(q',p)$ implies either $p\le p'\le 1/2$ or $p'\ge p\ge 1/2$. But in either case, the functions ϕ_1 and ϕ_2 are concave (in fact, affine) on the domain $[p,p']$ and hence $V(\alpha q+(1-\alpha)q',p)\le$ (in fact, $=$) $V(q,\alpha p+(1-\alpha)p')$, as desired. ■

The generalized utilitarian representation for $\triangle(\mathcal{I}\times\mathcal{X})$. We next show that we can

use essentially the same proof as for theorem 1 to obtain the form of generalized utilitarian representation given in expression (3) for an impartial observer's preferences $\succsim$ defined on $\triangle(\mathcal{I} \times \mathcal{X})$ that satisfy the axioms given in section 6. The key step is to show that the analog of lemma 7 (spanning) part (b) still applies: there exist two outcome lotteries ℓ^1 and ℓ_2 and four individuals i^1, i_1, i^2, and i_2 such that for any joint distribution $\left(z, (\ell_i)_{i\in\mathcal{I}}\right)$, either $\left(i^1, \ell^1\right) \succsim \left(z, (\ell_i)_{i\in\mathcal{I}}\right) \succsim \left(i_1, \ell^1\right)$ or $\left(i^2, \ell_2\right) \succsim \left(z, (\ell_i)_{i\in\mathcal{I}}\right) \succsim (i_2, \ell_2)$ or both. That is, we can still use two sets of *product* lotteries, one associated with ℓ^1 and one with ℓ_2, to span the entire range of the the impartial observer's preferences even though these are now defined over the full set of joint distributions $\triangle(\mathcal{I} \times \mathcal{X})$.

To see this, let $(\hat{z}, \left(\hat{\ell}_i\right)_{i\in\mathcal{I}})$ be an element of $\triangle(\mathcal{I} \times \mathcal{X})$ with the property that $(\hat{z}, \left(\hat{\ell}_i\right)_{i\in\mathcal{I}}) \succsim (z, (\ell_i)_{i\in\mathcal{I}})$ for all $(z, (\ell_i)_{i\in\mathcal{I}}) \in \triangle(\mathcal{I} \times \mathcal{X})$. By constrained independence, there must exist an individual i^1 in the support of $\hat{z}$ such that $(i^1, \left(\hat{\ell}_i\right)_{i\in\mathcal{I}}) \sim (\hat{z}, \left(\hat{\ell}_i\right)_{i\in\mathcal{I}})$. Let $\ell^1 := \hat{\ell}_{i^1}$, and let $\left(i^1, \ell^1\right)$ denote the (product) lottery $(i^1, (\ell_i)_{i\in\mathcal{I}})$ where $\ell_i = \ell^1$ for all $i \in \mathcal{I}$. By the acceptance* principle, $\left(i^1, \ell^1\right) \sim (i^1, \left(\hat{\ell}_i\right)_{i\in\mathcal{I}})$. Therefore, there exists an outcome lottery ℓ^1 and an individual i^1 such that the product lottery $\left(i^1, \ell^1\right)$ has the property that $\left(i^1, \ell^1\right) \succsim (z, (\ell_i)_{i\in\mathcal{I}})$ for all $(z, (\ell_i)_{i\in\mathcal{I}}) \in \triangle(\mathcal{I} \times \mathcal{X})$. Similarly, there exists a outcome lottery ℓ_2 and an individual i_2 such that the product lottery (i_2, ℓ_2) has the property that $(z, (\ell_i)_{i\in\mathcal{I}}) \succsim (i_2, \ell_2)$ for all $(z, (\ell_i)_{i\in\mathcal{I}}) \in \triangle(\mathcal{I} \times \mathcal{X})$. Define i_1 and i^2 exactly as in lemma 7. The proof of part (a) of lemma 7 (spanning) then follows with no change in the proof. And the analog of part (b) of the lemma (as stated above) follows immediately from part (a).

Thereafter, the proof of the representation result is almost unchanged. The analog of lemma 8 obtains a affine representation of the form $V(z, \ell) = \sum_{i=1}^{I} z_i V_i(\ell_i)$. The proof is the same as that for lemma 8 except that constrained independence is used wherever independence over identity lotteries was used before. This extends the representation from product lotteries $\triangle(\mathcal{I}) \times \triangle(\mathcal{X})$ to the full space of joint distributions $\triangle(\mathcal{I} \times \mathcal{X})$. The fact that $V_i(\ell_i)$ takes the form $\phi_i(U_i(\ell_i))$ follows (as before) from acceptance and outcome independence for individuals.

References

Arrow, Kenneth, J. (1963): *Social Choice and Individual Values*, 2nd edition, New York, Wiley.

Arrow, Kenneth, J. (1977): "Extended Sympathy and the Possibility of Social Choice", *American Economic Review*, 67, 219-225

Broome, John (1984): "Uncertainty and Fairness," *Economic Journal*, 94, 624-632.

Broome, John (1991): *Weighing Goods*, Oxford, Blackwell.

Broome, John (1993): "A cause of preference is not an object of preference," *Social Choice and Welfare*, 10, 57-68.

Diamond, Peter A. (1967): "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility: Comment." *Journal of Political Economy* 75, 765-66.

Blackorby, Charles, Walter Bossert and David Donaldson (2005): *Population Issues in Social Choice Theory, Welfare Economics and Ethics*, Cambridge, Cambridge University Press.

Blackorby, Charles, David Donaldson and Philippe Mongin (2004): "Social Aggregation without the Expected Utility Hypothesis," working paper 2004-020, Ecole Polytechnique.

Blackorby, Charles, David Donaldson and John Weymark (1999): "Harsanyi's social aggregation theorem for state-contingent alternatives", *Journal of Mathematical Economics*, 32, 365-387.

Elster, Jon (1989): *Solomonic Judgements: Studies in the Limitation of Rationality*, Cambridge: Cambridge University Press

Epstein, Larry G. and Uzi Segal (1992): "Quadratic Social Welfare," *Journal of Political Economy* 100, 691-712.

Ergin, Haluk and Faruk Gul (2009): "A Theory of Subjective Compound Lotteries," *Journal of Economic Theory*, 144(3), pp. 899-929.

Fishburn ,Peter C. (1982): *Foundations of Expected Utility*. Dordrecht: D. Reidel.

Fleurbaey, Marc (2007): "Assessing Risky Social Situations," mimeo University Paris, Descartes.

Grant, Simon, Atsushi Kajii, Ben Polak and Zvi Safra (2006): "Generalized utilitarianism and Harsanyi's impartial observer theorem", Cowles Foundation Discussion Papers number 1578.

Hammond, Peter (1981): "Ex-Ante and Ex-Post Welfare Optimality under Uncertainty ," *Economica*, 48, 235-250.

Hammond, Peter (1982): "Utilitarianism, Uncertainty and Information," in A.K. Sen and B. Williams (eds) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press, ch 4, 85 - 102.

Harsanyi, John C. (1953): "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking." *Journal of Political Economy* 61, 434-5.

Harsanyi, John C. (1955): "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility: Comment." *Journal of Political Economy* 63, 309-21.

Harsanyi, John C. (1975): "Nonlinear Social Welfare Functions: Do Welfare Economists Have a Special Exemption from Bayesian Rationality?" *Theory and Decision* 6, 311-32.

Harsanyi, John C. (1977): *Rational Behavior and Bargaining Equilibrium in Games and Social Situtations*, Cambridge, Cambridge University Press.

Karni, Edi and Zvi Safra (2000): "An extension of a theorem of von Neumann and Morgenstern with an application to social choice." *Journal of Mathematical Economics* 34, 315-27.

Karni, Edi and Zvi Safra (2002): "Individual sense of justice: a utility representation" *Econometrica* 70(1), 263-284.

Karni, Edi and John Weymark (1998): "An informationally parsimonious impartial observer theorem", *Social Choice and Welfare*, 15, 321-32.

Meyer, Margaret (1991): "A Social Welfare Function Approach to the Measurement of Ex Post Inequality under Uncertainty," working paper Nuffield College.

Mongin, Philippe (2001): "The Impartial Observer Theorem of Social Ethics," *Economics & Philosophy*, 17, 147-179.

Mongin, Philippe (2002): "Impartiality, Utilitarian Ethics and Collective Bayesianism", Laboratoire D'Econometrie, working paper 2002-030.

Mongin, Philippe and Claude d'Aspremont (1998): "Utility Theory and Ethics," in Salvador Barbera, Peter Hammond and Christian Seidl (eds), *Handbook of Utility Theory,* Boston, Kluwer Academic Publishers, chapter 8, pp 371-481.

Myerson, Roger (1981): "Utilitarianism, egalitarianism and the timing effect in social choice problems," *Econometrica* 49, 883-97.

Pattanaik, Prasanta K. (1968): "Risk, Impersonality, and the Social Welfare Function." *Journal of Political Economy* 76, 1152-69.

Rawls, John (1951): "Outline for a Decision Procedure for Ethics", *Philosophical Review*, 40, 177-197.

Safra, Zvi and Einat Weissengrin (2003): "Harsanyi's impartial observer theorem with a restricted domain." *Social Choice and Welfare*, 20(2), 95-111.

Sen, Amartya, K. (1970): *Collective Choice and Social Welfare.* San Francisco: Holden-Day.

Sen, Amartya, K. (1977): "On weights and measures: informational constraints in social welfare analysis." *Econometrica* 45, 1539-72.

Sen Amartya, K. (1979) "Interpersonal Comparisons of Welfare" in M. Boskin (ed.) *Economics and Human Welfare: Essays in Honor or Tibor Skitovsky*, New York: Academic Press, reprinted in Amartya, K. Sen (1982) *Choice, Welfare and Measurement*, Oxford: Blackwell, 264-282.

Vickrey, William (1945) "Measuring marginal utility by reaction to risk", *Econometrica*, 13, 319-33.

Weymark, John (1991): "A reconsideration of the Harsanyi-Sen debate on utilitarianism" in *Interpersonal Comparisons of Well-being* edited by Jon Elster and John Roemer. Cambridge: CUP, 255-320.