Human Capital and Economic Opportunity:
A Global Working Group

Working Paper Series

Working Paper No. 2012-012

Behavioral Fair Social Choice

Marc Fleurbaey
Erik Schokkaert

May, 2012

Human Capital and Economic Opportunity Working Group
Economic Research Center
University of Chicago
1126 E. 59th Street
Chicago IL 60637
humcap@uchicago.edu

Behavioral Fair Social Choice*

Marc Fleurbaey[†] Erik Schokkaert[‡]

April 3, 2012

Abstract

Behavioral economics has shaken the view that individuals have well-defined, consistent and stable preferences. This raises a challenge for welfare economics, which takes as a key postulate that individual preferences should be respected. This paper scrutinizes the challenge and argues, in agreement with Bernheim (2009) and Bernheim and Rangel (2009) that behavioral economics is compatible with consistency of partial preferences. While Bernheim and Rangel have focused on how to incorporate insights from behavioral economics into traditional concepts of welfare economics (Pareto optimality, compensation tests), we explore how the approach can be extended to deal with distributive issues. This paper revisits some key results of the theory in a framework with partial preferences and shows how one can derive partial orderings of individual and social situations.

Keywords: behavioral economics, preferences, welfare economics, psychology, social choice, fairness.

JEL Classification: D60, D71.

*This paper has benefited from detailed comments by François Maniquet and Giacomo Valletta, from discussions with D. Bernheim, A. Deaton, F. Dietrich, D. Kahneman, B. Tungodden, G. Weyl and from the reactions of audiences in Oxford, Maastricht and Leuven.

[†]Princeton University. Email: mfleurba@princeton.edu.

[‡]Department of Economics, Katholieke Universiteit Leuven, and CORE (Université catholique de Louvain). Email: erik.schokkaert@econ.kuleuven.be.

1 Introduction

One of the challenges for welfare economics is the formulation of adequate criteria to evaluate (re)distribution. Without such criteria policy evaluation can only be based on the Pareto criterion. Pareto-improving policy measures are rare, however. Rejecting all other policies leads to a conservative defence of the status quo, while the Kaldor-Hicks criterion of potential Pareto improvements is lacking ethical content. Indeed, the existence of a "potential improvement" is not very relevant if the necessary compensations remain purely hypothetical. To go beyond these Pareto-type approaches, one needs a concept of interpersonally comparable well-being. Traditional welfare economics has struggled for a long time with the issue of interpersonal comparisons. Arrow's impossibility theorem has most often been interpreted as showing that the informational basis of ordinal preferences is insufficient to derive an ordering of social states. In the wake of Sen (1970), a large literature explored the consequences of going beyond such ordinal preference information and derived welfare criteria under different assumptions about interpersonal comparability and measurability of individual subjective welfare (for an overview of this so-called welfarist approach, see d'Aspremont and Gevers, 2002).

The best way, even the possibility, to measure subjective welfare in an interpersonally comparable way remains, however, a controversial question. Fortunately, recent developments in the theory of fair allocations have shown that the common interpretation of Arrow's theorem is wrong and that an interpersonally comparable measure of subjective utility is not needed. According to these developments, fairness principles recommend to construct interpersonally comparable concepts of well-being that are actually based only on information about ordinal "non-comparable" individual preferences (for an overview, see Fleurbaey and Maniquet, 2011). One attractive approach is based on the concept of equivalent income, which is firmly rooted in the tradition of money-metric utility (Samuelson, 1974). This approach produces social criteria that respect individual preferences and are able to give some priority to the worse-off in the evaluation of public policies.

This so-called fairness approach offers a promising way out of Arrow's impossibility without necessitating the use of subjective utilities, but it does rest on the assumption that well-defined individual preferences exist. The findings of behavioral economics have cast doubt on this assumption. The existence of "behavioral anomalies" suggests that it is difficult to interpret individual choice behavior as the maximization of well-defined preferences. This has important implications for welfare economics. Some authors (Frey and Stutzer, 2002; Kahneman et al., 1997; Kahneman and Sugden, 2005; Köszegi and Rabin, 2008; Layard, 2005) have advocated to focus on experience utility (and subjective happiness) rather than on decision utility. This would bring us back to the welfarist interpretation of Arrow's theorem. Other authors refuse to take this step and formulate preference- or choice-based welfare criteria (Bernheim and Rangel, 2009; Bernheim, 2009; Dalton and Ghosal, 2010; Rubinstein and Salant, 2009; Salant and Rubinstein, 2008). The proposed preference relations are incomplete and the question remains whether it is possible in this approach to go beyond the Pareto-criterion.

In this paper we bring together these two recent streams of literature. We examine if it is still possible to derive an interpersonally comparable concept of well-being and a tractable criterion for the evaluation of policies, when one works with an incomplete preference relation as defined, e.g., by Bernheim and Rangel (2009). We show that the answer to this question is positive. Using the incomplete individual preference relation proposed by Bernheim and Rangel (2009), we derive an incomplete ordering of personal situations in terms of well-being and we argue that this concept of well-being, which relies only on ordinal preferences, can be used for distributional judgments. Respect for individual preferences is the key value in our approach and we explore how far one can go if one accepts this key value.

We briefly recall in Section 2 why behavioral welfare economics threatens approaches that involve standard individual preferences, including a social welfare approach that would invoke "authentic" preferences as the yardstick of well-being. Sections 3 and 4 show how a theory of fair social choice, relative to interpersonal comparisons (Sect. 3) and social evaluation (Sect. 4), can be developed for the case of incomplete preferences. Section 5 concludes.

2 Behavioral economics: shaking preferences?

It is not our point here to give a complete overview of all so-called behavioral anomalies that have been described in the literature, as there exist by now a lot of survey papers. Referring to just one of these that focuses on evidence from the field (Della Vigna, 2009), one can distinguish *non-standard preferences (*self-control problems in an intertemporal setting; the influence of default options and the endowment effect), *non-standard beliefs* (economic agents overestimate their performance in tasks requiring ability, they expect small samples to exhibit large-sample statistical properties and they project their current preferences onto future periods) and *non-standard decision making* (the neglect or overweighting of information because of limited attention; the use of suboptimal heuristics for choices out of menu sets; excess impact of the beliefs of others; the possibly important role played by emotions such as mood and arousal). The findings from this literature suggest that preferences may not be well-behaved – and that, even if standard preferences did exist, choice behavior cannot in any case be interpreted as the simple maximization of a fixed preference ordering. This raises difficult challenges for welfare economics.

One popular reaction in the behavioral literature has been to go back to experience utility (Frey and Stutzer, 2002; Kahneman et al., 1997; Kahneman and Sugden, 2005; Köszegi and Rabin, 2008; Layard, 2005). The intuition behind this is that, if people make mistakes, "decision utility" (the perceived utility on which decisions are based) and "experience utility" (the real after-decision utility) do no longer coincide and that in these circumstances it is better from the welfare point of view to focus on the "correct" outcomes. Yet, this move back to welfarism is a very controversial approach. In particular subjective utility is not directly comparable across individuals or even for a same individual at different dates, when the levels of utility to be compared involve different standards of evaluation. For instance, the subjective satisfaction of a given population may appear stable over time in spite of their judging that their situation has greatly improved, just because their standards of evaluation evolve with their situation, a phenomenon known as adaptation (or as the aspiration treadmill). Therefore ranking individual situations on the basis of happiness does not respect individual preferences.[1]

The alternative approach is to keep preferences as the ultimate criterion for evaluating social states, but to take into account that the preference relation that can be derived from behavior is not standard if choices (or stated preferences)[2] are conflicting and context-dependent. An interesting way to model context-dependency has been proposed by Bernheim and Rangel (2009) and Salant and Rubinstein (2008).[3] They introduce the concept of a generalized choice situation (A, d) where A is the set of elements from which a choice has to be made and d is an "ancillary condition" (in the terminology of Bernheim and Rangel, 2009) or a "frame" (in the terminology of Salant and Rubinstein, 2008). A standard choice situation would be fully characterized by A. Ancillary conditions (or frames) influence decisions but are (by definition) irrelevant for welfare. Examples of frames could be the specification of a default option or circumstances which lead to emotional arousal.[4] The set of all generalized choice situations of interest is given by $\mathbb{C}$. The choice-correspondence for individual i is then given by $C_i(A, d) \subseteq A$ for all $(A, d) \in \mathbb{C}$. Its interpretation is obvious: $x \in C_i(A, d)$ is an object that individual i may choose when facing (A, d). Note that it is very well possible that $C_i(A, d) \neq C_i(A, d')$. This is precisely

[1] Fleurbaey et al. (2009) develop this point and argue that happiness data can still be very useful to recover information about individual ordinal preferences. See also Bernheim (2009); Loewenstein and Ubel (2008).

[2] Following the literature, we focus on choice in this part of the paper, but there is no reason to ignore other sources of data on preferences, such as stated preferences.

[3] Another interesting framework in terms of binary relations over swaps is proposed by Gustafsson (2011).

[4] Bernheim and Rangel (2009) and Salant and Rubinstein (2008) give many examples on how to cast the behavioral anomalies from the literature in the mould of generalized choice situations.

how behavioral "anomalies" can be integrated in this framework.

If one has observations of individual behavior in different generalized choice situations, one can derive information about individual preferences over A. One possibility is to apply a structural model of behavior to explain the observations $C_i(A,d)$, i.e., to model how preferences together with frames determine choice. This structural model can then be used to derive a preference relation that is consistent with observed behavior conditional on the model used (Dalton and Ghosal, 2010; Rubinstein and Salant, 2009). The revealed preference relation is not necessarily complete. More importantly, it will depend on the specific behavioral model that is applied – and, very often it will be impossible to identify the correct model from the observations, in the sense that the outcomes of two different behavioral models (with different underlying preference relations) are observationally equivalent (in terms of choices).

Bernheim and Rangel (2009) therefore propose as an alternative what they call a "libertarian" approach, because it only uses information about choices.[5] On this basis they define a series of incomplete welfare relations. The most attractive (and the one with which they themselves work extensively) is

$$xP_i^*y \text{ iff for all } (A,d) \in \mathbb{C} \text{ such that } x,y \in A, \text{ one has } y \notin C_i(A,d).^6$$

Bernheim and Rangel (2009) have a counterexample showing that P_i^* is not necessarily transitive, but they show that P_i^* is acyclic. Imposing more structure on the space of alternatives may lead to P_i^* being transitive and, in fact, for almost all popular behavioral approaches, P_i^* is indeed transitive.

Bernheim and Rangel (2009) emphasize that their approach is only choice-based and

[5] As already explained, one could also use additional information coming from stated preferences. The main point is to avoid making structural assumptions about how choice or preference behavior is determined.

[6] One can add the condition that there is at least one $(A,d) \in \mathbb{C}$ such that $x,y \in A$ for which $x \in C_i(A,d)$. This condition is always satisfied in Bernheim and Rangel (2009), because they assume that $(\{x,y\},d) \in \mathbb{C}$ and the individual always selects some alternative in each generalized choice situation.

does not assume the existence of an underlying preference relation. But their formalism is compatible with a variety of interpretations. As a matter of fact, they are well aware that, in a purely choice-based approach the relation P^* can be very coarse – and that in some generalized choice situations it is highly unlikely that a choice reveals something about welfare. The foreigner who is killed in a car accident in London because he did forget to look right when crossing the road does not reveal that he preferred being killed (unless there is other evidence in his life suggesting that he had suicidal tendencies). They therefore consider the possibility of refining P^* by using non-choice information to discard information from some generalized choice situations as "suspect". The decisions of what are "suspect" generalized choice situations will necessarily be to some extent subjective. Clear cases of biased information and misperception are perhaps easy to judge. Criteria of simplicity and coherence between behavior in different situations are already more debatable – and the use of self-reported preferences or informed expert opinions (as advocated, e.g., by Beshears et al., 2008) is largely rejected by Bernheim (2009).

For applied welfare analysis, Bernheim and Rangel (2009) introduce natural counterparts of the concepts of compensating and equivalent variation. We will focus on the former. Let us assume that the generalized choice situation can be written as $((A(\alpha, m), d)$, where α is a vector of environmental parameters and m is a monetary transfer. Let us then consider a move from $((A(\alpha_0, 0), d_0)$ to $((A(\alpha_1, m), d_1)$. The compensating variation is the smallest value of m such that for any $x \in C(A(\alpha_0, 0), d_0)$ and $y \in C(A(\alpha_1, m), d_1))$ the individual would be willing to choose y over x. In a setting with incomplete preferences, the latter sentence is ambiguous, however. We can consider the compensation to be sufficient when the new situation is unambiguously chosen over the old one or when the old situation is not unambiguously chosen over the new one. This leads to two notions of compensating variation. The first, CV^{high}, is equal to

$$\inf \left\{ m | yP^*x \text{ for all } m' \geqslant m, x \in C(A(\alpha_0, 0), d_0) \text{ and } y \in C(A(\alpha_1, m'), d_1) \right\}$$

The second, CV^{low}, is equal to

$$\sup\left\{m|xP^{*}y \text{ for all } m' \leqslant m, x \in C(A(\alpha_0, 0), d_0) \text{ and } y \in C(A(\alpha_1, m'), d_1)\right\}$$

It is easy to see that $CV^{high} \geq CV^{low}$.

In a setting with several individuals, a move from x to y is a Pareto-improvement if yP_i^*x for all i. If we do not define an interpersonally comparable concept of well-being, policy analysis remains restricted to looking for such Pareto-improvements. These will be very rare indeed and this usually motivates the use of the sum of compensating variations. While the compensating variation yields a specific measure of the welfare change for one individual, it is well known, however, that simply adding compensating variations is not an acceptable welfare criterion if one cares about the distribution and if one wants to avoid cyclic decisions (Blackorby and Donaldson, 1990). A setting with incomplete preferences does nothing to alleviate this criticism. This is why better measures of well-being, which allow for interpersonal comparisons, are needed, such as those studied in the next sections. Some of these measures correspond to particular variants of money-metric utility, and the idea of upper and lower bounds just presented will appear again as the natural way to cope with incomplete preferences.

Before examining these alternative measures, a preliminary question remains to be addressed. One may wonder whether behavioral economics is really a challenge for the normative perspective taken by the theory of social choice. The specific individual preferences one has in mind in this normative context are people's "authentic preferences" about what a good life is – not the preferences that are revealed in actual choice behavior. Authentic preferences are the preferences that would be revealed, under "perfect" (or sufficiently good) circumstances, by well-informed individuals who consider all the relevant aspects of life and are not misled by cues of the environment. It could be argued that these preferences must necessarily be complete. If individuals are influenced by the framing of the choice problem or use imperfect information, this should be corrected. In the terminology introduced before, there is much scope and justification, in this context, for refinements. Note that the argumentation about the refinements here can take an ethical twist: some ancillary conditions can be discarded even if they do not unambiguously involve biases, just because they are ethically unappealing. One example could be the importance of the reference situation – it is well known that people tend to focus on changes (gains and losses) rather than on the resulting final states and, moreover, that the feeling of loss looms larger than the feeling of gain. People will give larger subjective weight to avoiding the former than to experiencing the latter. One could argue that from the ethical point of view the status quo position should play a less prominent role, and definitely so if one is concerned about evaluating redistribution measures that are to the advantage of the poorest in society, i.e. where the rich lose and the poor gain. Of course, a justification of this kind of position ultimately rests on an ethical argumentation, that will not necessarily be accepted by everyone.[7]

This position that "authentic preferences" must necessarily be complete is too easy, however. First, even if one accepts that such authentic preferences do exist, there is still a challenge to discover them. Whether we simply ask people about their ideas, or we let them participate in choice experiments, or we start from real-life choices, or we ask them about their overall satisfaction with life – these "revealed" and "stated" preferences will be partly influenced by the way the choice situation is set up. Emotions and heuristics will also play a role in evaluating one's overall life. It is possible that in a cold-blooded evaluation of the life cycle, emotions may receive less weight than in the heat of the moment. Yet, even if individuals are invited to abstract from immediate emotions, and even if they take a cognitive stance, they will still follow reason-based heuristics and the context will have an impact on their decisions (Shafir et al., 1993; Tversky and Simonson,

[7]Noor (2010) suggests that choice-theoretic foundations for normative preferences can be built on the idea that the normative perspective is revealed when the agent is distanced from the consequences of his choice, an idea which is closely related to the veil of ignorance.

1993). Therefore, the concept of generalized choice situation is relevant also in this context – and the analyst will only be able to "derive" an incomplete (preference) relation P^*. While refinements are indeed possible, it is unlikely that there will be complete consensus about them or that we can refine until we ultimately end up with standard preferences. Moreover, the argument of observational equivalence between different behavioral models remains relevant in this setting too. Therefore, even if one accepts that individuals have well-defined authentic preferences, any cautious application will still have to admit our lack of knowledge about them. Working with incomplete preferences can then be seen as a kind of robustness check.

Second, one can even doubt that individuals have a complete preference relation over all possible lives. Indeed, this would imply that they can order states with which they are not at all familiar. The psychological uncertainty about preferences may be expected to grow when one goes further away from the actual situation. Yet, to calculate, e.g., money-metric utilities,[8] we need information on the whole indifference curve. Is someone who has been chronically ill for a long time (or is handicapped since birth) able to evaluate trade-offs in a situation of (nearly) perfect health? Note that we are referring here to ordinal preferences and not to the effect of adaptation leading to smaller changes in subjective satisfaction levels.

The question of the existence of a "true" underlying well-defined and complete preference ordering is interesting from a philosophical point of view, but at this stage we can leave it open. For our purposes the first argument is already sufficient: in most cases, the analyst (or the policy-maker) will have to work with incomplete preferences. The formal framework that was described earlier will therefore be relevant. However, the economic models that will be studied in the next sections put more structure on the decision problem than the abstract and general approach of Bernheim and Rangel (2009). Let us therefore conclude this section by describing the form taken by preference relations in our approach.

[8] See Section 4 for a formal definition of this notion.

We assume that individual preferences take the form of *partial* binary relations P^* defined on the set of relevant life dimensions X, with $X = \mathbb{R}_+^\ell$. Examples of life dimensions could be consumption, health status, job satisfaction, quality of interpersonal relations, etc. The expression xP^*y means that x is strictly preferred to y. We assume that P^* is transitive (xP^*y and yP^*z implies xP^*z) and irreflexive (xP^*x for no $x \in X$). As noted before, transitivity is not a very strong requirement. In our setting it also makes sense – certainly as a first approach – to assume that preferences are monotonic ($x > y$ implies xP^*y)[9] and continuous. We define continuity as meaning that the sets

$$UC(x, P^*) = \{q \in X \mid qP^*x\}$$

and

$$LC(x, P^*) = \{q \in X \mid xP^*q\}$$

are open subsets of X, and in addition $x \in \partial UC(y, P^*)$ if and only if $y \in \partial LC(x, P^*)$, where $\partial UC(.)$ denotes the lower boundary of $UC(.)$ and $\partial LC(.)$ the upper boundary of $LC(.)$. Let

$$NC(x, P^*) = \{q \in X \mid \text{neither } qP^*x \text{ nor } xP^*q\}$$

be the set of vectors which are not comparable to x by P^*.[10]

Under monotonicity and transitivity, $UC(x, P^*)$ is upper comprehensive, i.e., if $q \in UC(x, P^*)$, and $q' > q$, then $q' \in UC(x, P^*)$. Similarly, $LC(x, P^*)$ is lower comprehensive, i.e., if $q \in LC(x, P^*)$ and $q' < q$, then $q' \in LC(x, P^*)$.

Under these conditions, it is therefore enough to know $NC(x, P^*)$ in order to know

[9] Vector inequalities are denoted $\geq, >, \gg$.

[10] Note that only assuming that $UC(x, P^*)$ and $LC(x, P^*)$ are open would not prevent some forms of discontinuity, namely, the appearance of "poles" at some points: one could have a sequence $x_n \to x$ such that $y \in LC(x_n, P^*)$ for all n but y belongs to the interior of $X \setminus LC(x, P^*)$. Under the additional condition that $x \in \partial UC(y, P^*)$ if and only if $y \in \partial LC(x, P^*)$, such phenomenon cannot occur because x is in the interior of $NC(y, P^*)$ if and only if y is in the interior of $NC(x, P^*)$. Indeed, assume that x is in the interior of $NC(y, P^*)$. Then one cannot have $y \in \partial NC(x, P^*)$, which would require either $y \in \partial UC(x, P^*)$ or $y \in \partial LC(x, P^*)$. And one cannot have $y \notin NC(x, P^*)$, which would require either $y \in UC(x, P^*)$ or $y \in LC(x, P^*)$.

$UC(x,P^*)$ and $LC(x,P^*)$. One has

$$\begin{aligned} UC(x,P^*) &= \left\{q \in X \mid q \notin NC(x,P^*) \text{ and } \exists q' \in NC(x,P^*),\ q > q'\right\} \\ LC(x,P^*) &= \left\{q \in X \mid q \notin NC(x,P^*) \text{ and } \exists q' \in NC(x,P^*),\ q < q'\right\}. \end{aligned}$$

If xP^*y, then $LC(y,P^*) \subsetneq LC(x,P^*)$, as we now show. First, one cannot have $LC(x,P^*) = LC(y,P^*)$ because $y \in LC(x,P^*)$ but $y \notin LC(y,P^*)$. Second, suppose that $LC(y,P^*) \subseteq LC(x,P^*)$ does not hold. Let $z \in LC(y,P^*) \setminus LC(x,P^*)$. One has yP^*z, which by transitivity implies xP^*z and therefore $z \in LC(x,P^*)$, a contradiction. Similarly, one shows that if xP^*y, then $UC(x,P^*) \subsetneq UC(y,P^*)$.

In the next sections it will be useful to address the following question. Consider two preferences $P^*, P^{*\prime}$ and two sets of points Q, Q'. Under what conditions does there exist a preference $P^{*\prime\prime}$ such that for all $q \in Q$, $NC(q,P^{*\prime\prime}) = NC(q,P^*)$ and for all $q' \in Q'$, $NC(q',P^{*\prime\prime}) = NC(q',P^{*\prime})$? A sufficient condition is that for all $q \in Q$, $q' \in Q'$, $NC(q,P^*) \cap NC(q',P^{*\prime}) = \varnothing$. This condition is an extension of the notion of non-crossing indifference curves.

The resulting indifference curves in the two-dimensional case are represented in Figure 1. For convenience, we have drawn them in a strictly convex way, but convexity is not necessary for our analysis. Note that our assumptions imply that individuals have finer preferences when comparing close alternatives: this seems a very natural assumption.

3 Interpersonal comparisons with incomplete preferences

Let us first consider the issue of interpersonal comparisons, i.e., of ranking personal situations (x,P^*) in terms of well-being. The object to be constructed is a binary relation on such situations, that is denoted $\succcurlyeq$ (with asymmetric and symmetric parts $\succ, \sim$) and is required to be reflexive and transitive (i.e., it is a preordering) but not necessarily complete.

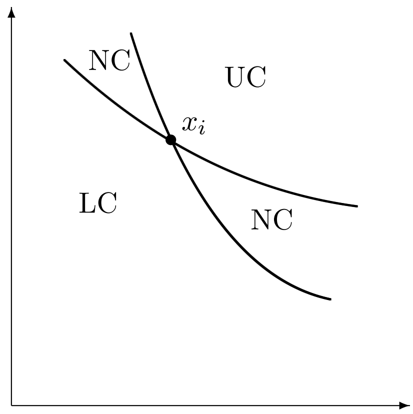

Figure 1: Indifference curves for incomplete preferences

The model is the same as in the previous section. We extend the characterization of the equivalence approach (to be defined below) that is provided by Fleurbaey et al. (2009). To simplify the analysis, anonymity is assumed from the outset, so that the identity of individuals is not part of the description of situations (x, P^*). The axioms and concepts from Fleurbaey et al. (2009) can be adapted in a straightforward way. Universal quantifiers are omitted whenever the meaning of the axiom is clear.

First, the preference principle says that when two situations share the same preferences, the preordering should agree with them. This is the core principle that guarantees respect for individual preferences. Note that it applies not just for intra-personal comparisons, but also for interpersonal comparisons. Note that this preference principle is incompatible with ranking the individual life situations on the basis of subjective well-being (or happiness): it is very well possible that two individuals agree that x is better than x', and that at the same time the individual in situation x has a lower level of subjective happiness than the individual in situation x', e.g., because she has more ambitious aspirations.

Preference Principle: $(x, P^*) \succ (x', P^*)$ if xP^*x'.

Second, the dominance principle says that when a bundle x dominates another, the corresponding situation is preferable independently of the associated preferences.

Dominance Principle: $(x, P^*) \succcurlyeq (x', P^{*\prime})$ if $x \geq x'$; $(x, P^*) \succ (x', P^{*\prime})$ if $x \gg x'$.

However, it is well known that in the case of complete individual preferences, the preference principle and the dominance principle are incompatible. This obviously extends to incomplete preferences because the latter principle implies that $(x, P^*) \sim (x, P^{*\prime})$ for all x and all $P^*, P^{*\prime}$, making it impossible to take account of preferences. Even the second part of the dominance principle it by itself incompatible with the preference principle. This is shown by the following example from Brun and Tungodden (2004). Assume $X = \mathbb{R}^2_+$ and take $x_i, x_j, x'_i, x'_j \in X$ and P^*_i, P^*_j such that $x_i \gg x_j, x'_i \ll x'_j$, $x'_i P^*_i x_i$, and $x_j P^*_j x'_j$. Figure 2 illustrates this configuration. The preference principle implies that (x'_i, P^*_i) is better than (x_i, P^*_i) and (x_j, P^*_j) is better than (x'_j, P^*_j) while the dominance principle implies that (x_i, P^*_i) is better than (x_j, P^*_j) and (x'_j, P^*_j) is better than (x'_i, P^*_i). By transitivity, one obtains that (x_i, P^*_i) is better than (x_i, P^*_i), which is impossible.

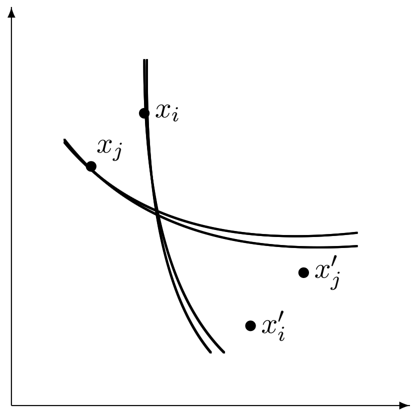

Figure 2: Incompatibility of the preference and the dominance principles

In order to cope with this incompatibility, Fleurbaey et al. (2009) restrict the dominance principle to a subset B of X:

Restricted Dominance Principle: For all $x, x' \in B$, $(x, P^*) \succcurlyeq (x', P^{*\prime})$ if $x \geq x'$; $(x, P^*) \succ (x', P^{*\prime})$ if $x \gg x'$.

The *equivalence approach* gathers the preorderings $\succcurlyeq$ such that how to rank (x, P^*) and $(x', P^{*\prime})$ is fully determined by $NC(x, P^*) \cap B$ and $NC(x', P^{*\prime}) \cap B$. In the case of complete preferences, the intersection of B with indifference sets is the relevant information for this approach. As explained in Fleurbaey (2009), the equivalence approach is not new. It basically boils down to the idea of money-metric utility that was popular in the 1980s (Samelson, 1974; Deaton and Muellbauer, 1980; Willig, 1981; King, 1983). In the current framework, one obtains the following result. It refers to the notion of a monotone path, which is defined as a set $A \subseteq X$ such that $0 \in A$, A is unbounded and connected, and for all $x, x' \in A$, either $x \geq x'$ or $x \leq x'$.

Proposition 1 *Let B be a subset of X such that for all (x, P^*), $NC(x, P^*) \cap B \neq \varnothing$. If $\succcurlyeq$ satisfies the preference principle and the restricted dominance principle with respect to B, then B is a monotone path and $(x, P^*) \succ (x', P^{*\prime})$ whenever $LC(x, P^*) \cap UC(x', P^{*\prime}) \cap B \neq \varnothing$.*

Proposition 1 does allow for different ways of making interpersonal comparisons, since it does not fix the monotone path (just as money-metric utilities depend on the choice of reference prices). However, the recent literature on fair allocations has shown that there may be good ethical reasons to choose a specific monotone path (Fleurbaey and Maniquet, 2011). Although some open questions remain, attractive solutions have been found for specific policy environments. One example (for health) will be described in the next section.

More specifically to our setting, Proposition 1 does not characterize the ranking fully. Assuming that B is a monotone path, every ranking such that:

(i) $(x, P^*) \succ (x', P^{*\prime})$ whenever $LC(x, P^*) \cap UC(x', P^{*\prime}) \cap B \neq \varnothing$,

(ii) $(x, P^*) \sim (x, P^{*\prime})$ whenever $x \in B$,

(iii) $(x, P^*) \succ (x', P^*)$ whenever xP^*x',

satisfies the preference principle and the restricted dominance principle. This allows many different possible rankings when none of these three situations applies. In particular, such rankings may involve the equivalence approach w.r.t. other paths, or non-equivalence approaches. As an example, compare the three cases in Figures 3-5. Proposition 3 allows us to say that $(x, P^*) \succ (x', P^{*\prime})$ in case A – but it does not allow a similar conclusion in cases B and C.

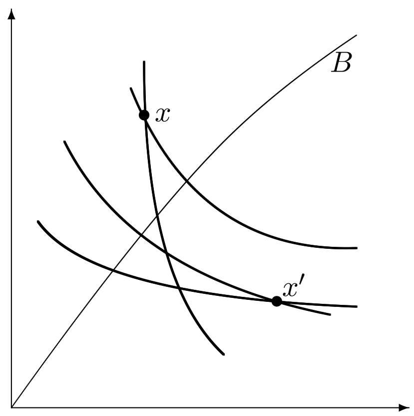

Figure 3: Case A

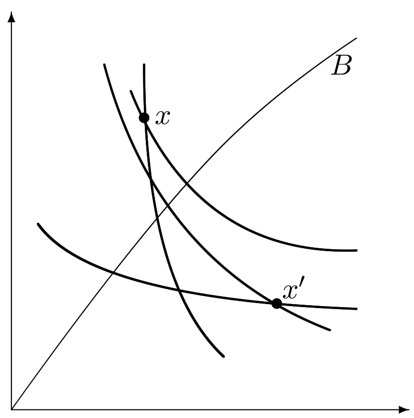

Figure 4: Case B

We get closer to the equivalence approach by adding additional requirements. One attractive condition is that one should avoid ranking an individual as better off than another individual when the available information about his situation is compatible with his being unambiguously worse off. This intuition is captured by the following Safety

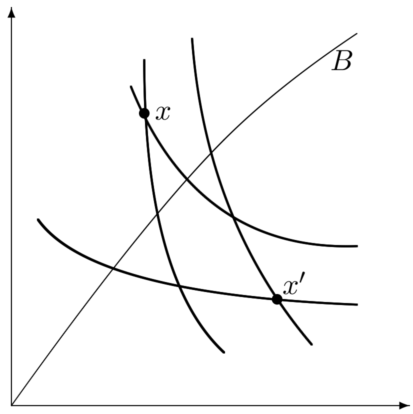

Figure 5: Case C

Principle:

Safety Principle: $(x, P^*) \succ (x', P^{*\prime})$ if there exists $\overline{P^*}' \supseteq P^{*\prime}$ such that for all $\overline{P^*} \supseteq P^*$, $(x, \overline{P^*}) \succ (x', \overline{P^*}')$.

Specifically, the safety principle says that if a refinement of one individual's preferences may reveal him to be worse off, then we should already consider him to be worse off. Of course, there are situations in which either individual can turn out to be worse off than the other when the information about both agents is refined. But the axiom deals with the case in which refining the information about one of them only may already determine that he is worse off. The main motivation for this axiom is that, even though it does not preclude mistakes in interpersonal comparisons, it prevents the evaluator from missing a situation in which the worse-off is really badly off. If the evaluator is wrong about the worse-off in a pairwise comparison, the true worse-off is not as badly off as he could be if the mistake was in the opposite direction. Imposing it leads to the following proposition.

Proposition 2 *Let B be a subset of X such that for all (x, P^*), $NC(x, P^*) \cap B \neq \varnothing$. If $\succcurlyeq$ satisfies the preference principle and the restricted dominance principle with respect to B, then B is a monotone path and, under the safety principle, $(x, P^*) \succ (x', P^{*\prime})$ whenever $LC(x, P^*) \cap (X \setminus LC(x', P^{*\prime})) \cap B \neq \varnothing$.*

Returning to the examples in the Figures, application of the safety principle now makes it possible to state that $(x, P^*) \succ (x', P^{*\prime})$ also in cases B and C.

However, case C illustrates a limitation of the safety principle. Indeed, refining $P^{*\prime}$ to $\overline{P^*}'$ could also lead to the opposite configuration – implying that $(x, \overline{P^*}) \prec (x', \overline{P^*}')$. A variant of the safety principle is more cautious and checks that this cannot occur. Imposing it means that we can still draw conclusions in case B, but no longer in case C.

Super Safety Principle: $(x, P^*) \succ (x', P^{*\prime})$ if:

(i) there exists $\bar{P}^{*\prime} \supseteq P^{*\prime}$ such that for all $\bar{P}^* \supseteq P^*$, $(x, \bar{P}^*) \succ (x', \bar{P}^{*\prime})$ or there exists $\bar{P}^* \supseteq P^*$ such that for all $\bar{P}^{*\prime} \supseteq P^{*\prime}$, $(x, \bar{P}^*) \succ (x', \bar{P}^{*\prime})$;

(ii) there exists no $\bar{P}^{*\prime} \supseteq P^{*\prime}$ such that for all $\bar{P}^* \supseteq P^*$, $(x, \bar{P}^*) \prec (x', \bar{P}^{*\prime})$ and there exists no $\bar{P}^* \supseteq P^*$ such that for all $\bar{P}^{*\prime} \supseteq P^{*\prime}$, $(x, \bar{P}^*) \prec (x', \bar{P}^{*\prime})$.

Imposing Super Safety yields the following proposition.

Proposition 3 *Let B be a subset of X such that for all (x, P^*), $NC(x, P^*) \cap B \neq \varnothing$. If $\succcurlyeq$ satisfies the preference principle and the restricted dominance principle with respect to B, then B is a monotone path and, under the super safety principle, $(x, P^*) \succ (x', P^{*\prime})$ whenever $LC(x', P^{*\prime}) \cap (X \setminus LC(x, P^*)) \cap B = \varnothing$, $UC(x, P^*) \cap (X \setminus UC(x', P^{*\prime})) \cap B = \varnothing$, and either $LC(x, P^*) \cap (X \setminus LC(x', P^{*\prime})) \cap B \neq \varnothing$ or $UC(x', P^{*\prime}) \cap (X \setminus UC(x, P^*)) \cap B \neq \varnothing$.*

The formal statement of the following proposition is not very reader-friendly, but the meaning of the final conditions is simple: on B, the boundaries of $UC(x, P^*)$ and $LC(x, P^*)$ are not below the corresponding boundaries for $(x', P^{*\prime})$, and at least one of them is strictly above. This is illustrated in case C.[11]

It is clear that the propositions do not lead to a complete ranking of all possible personal situations (x, P^*). Yet, in many cases they give an attractive operational criterion

[11] In Case C strict inequality is observed for both boundaries – in the proposition one equality would be admitted, which extends the power of the preordering $\succcurlyeq$ a little.

to define who is the worst off in a pairwise comparison. Moreover, our approach shares an attractive feature with the Bernheim-Rangel (2009) framework: if preferences are refined we come closer and closer to the standard approach defined for complete preferences.

4 Social evaluation with incomplete preferences

Let us now turn to the issue of comparing social states. To obtain an intuitive identification of the monotone path, we work with a specific model in which the two relevant life dimensions are health and consumption (we follow Fleurbaey, 2005). An individual situation is $x_i = (c_i, h_i) \in X$, with $X = \mathbb{R}_+ \times [0, 1]$. The fixed population is $N = \{1, ..., n\}$ and an allocation is denoted $x_N = (x_1, ..., x_n)$. The incomplete individual preferences are denoted P_i^* and are assumed to be monotonic, transitive, irreflexive, and to satisfy the continuity property introduced at the end of Section 2.

The ranking of allocations from the point of view of social welfare will be denoted $\mathbf{R}$ (with asymmetric and symmetric parts $\mathbf{P}$ and $\mathbf{I}$), and will be assumed to be reflexive and transitive but not necessarily complete. Since we want this social ranking to depend on the profile of individual preferences $P_N^* = (P_1^*, ..., P_n^*)$, it is really a function $\mathbf{R}(P_N^*)$, but the argument will often be dropped to shorten notation.

The notion of equivalent income will play an essential role in what follows, and corresponds to a special case of the equivalence approach. It is obtained by choosing the monotone path B as the set of all points in X with $h = 1$.[12] It will become clear later in this section why this is a relevant choice.

In the case of complete individual preferences, the *healthy-equivalent income* is the quantity c^* implicitly defined by the condition

$$(c^*, 1) I_i (c_i, h_i), \tag{1}$$

[12] This set does not include 0, and one can add the bundles such that $c = 0$ to have a full path. But the result recalled below only deals with bundles that are at least as good as $(0, 1)$ for every individual.

where I_i is the indifference relation. In other words, c^* is the level of consumption that, combined with perfect health, would make the individual indifferent with his current situation (c_i, h_i). The concept is illustrated (for complete preferences) in Figure 6.

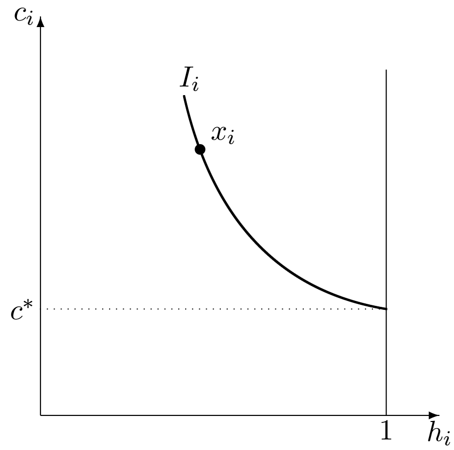

Figure 6: The healthy-equivalent income in the case of complete preferences

When preferences are incomplete, upper and lower bounds extend this notion in a natural way. For a given $x \in X$ and a given P^*, one can define the upper and the lower healthy-equivalent incomes

$$\begin{aligned} E^{\sup}(x, P^*) &= c^* \text{ such that } (c^*, 1) \in \partial UC(x, P^*), \\ E^{\inf}(x, P^*) &= c^* \text{ such that } (c^*, 1) \in \partial LC(x, P^*). \end{aligned}$$

We restrict attention to allocations x_N such that for all i, $E^{\inf}(x_i, P_i^*)$ and $E^{\sup}(x_i, P_i^*)$ are well defined.[13]

Note the close relationship with the concepts of compensating variation CV^{high} and CV^{low}, as proposed by Bernheim and Rangel (2009). However, healthy-equivalent incomes

[13] Note that the previous propositions can be reformulated for the special case of healthy-equivalent incomes. Indeed, for this specific choice of monotone path, we can derive the simple operational criteria:

(preference principle, restricted dominance principle) $(x, P^*) \succ (x', P^{*\prime})$ whenever $E^{\inf}(x, P^*) > E^{\sup}(x', P^{*\prime})$

(preference principle, restricted dominance principle, safety principle) $(x, P^*) \succ (x', P^{*\prime})$ whenever $E^{\inf}(x, P^*) > E^{\inf}(x', P^{*\prime})$

(preference principle, restricted dominance principle, super safety principle) $(x, P^*) \succ (x', P^{*\prime})$ whenever $E^{\inf}(x, P^*) \geq E^{\inf}(x', P^{*\prime})$ and $E^{\sup}(x, P^*) \geq E^{\sup}(x', P^{*\prime})$,with at least one strict inequality.

yield an interpersonally comparable measure of well-being, i.e., an evaluation of the individual's overall personal situation, and not only a monetary evaluation of a change in this personal situation. They can therefore be used for social evaluation in cases where the distribution matters.

In the standard framework with complete preferences, Fleurbaey (2005) formulates three requirements that can be imposed on the social preordering. They are easily adapted to the case of incomplete preferences. As in the previous section, universal quantifiers are omitted whenever the meaning of the axiom is clear.

Weak Pareto: If for all i, $x_i P_i^* x_i'$, then $x_N \mathbf{P}(P_N^*) x_N'$.

Independence: If for all i, $NC(x_i, P_i^*) = NC(x_i, P_i^{*\prime})$ and $NC(x_i', P_i^*) = NC(x_i', P_i^{*\prime})$, then $x_N \mathbf{R}(P_N^*) x_N'$ if and only if $x_N \mathbf{R}(P_N^{*\prime}) x_N'$.

Pigou-Dalton: If there is i, j such that $h_i = h_j$ and $(c_i, h_i) = (c_i' - \delta, h_i') > \left(c_j' + \delta, h_j'\right) = (c_j, h_j)$ for some $\delta > 0$ while $x_k' = x_k$ for all $k \neq i, j$, then $x_N \mathbf{R}(P_N^*) x_N'$ provided that either $P_i^* = P_j^*$ or $h_i = h_j = 1$.

The first axiom is standard. The second one defines the informational setting: it states that the social ranking of two allocations x and x' should be based only on information concerning the sets $NC(x_i, P_i^*)$ for all individuals i.[14] This allows for richer information than Arrow's impossibility theorem, which would only consider individual pairwise preferences over x and x'.

The third axiom introduces some egalitarianism in the space of resources. At first sight one could think that it would make sense to impose the restriction that a transfer of consumption from the rich to the poor increases social welfare, under the condition that the rich and the poor are at the same health level. However, Fleurbaey and Trannoy (2003) have shown that this requirement is incompatible with the Pareto condition. The

[14] In a setting with complete preferences the corresponding assumption would be that the social ranking is based on information concerning the shape of the indifference curves through x_i and x_i' for all individuals.

third condition above therefore restricts the application of resource transfers to individuals with the same preferences or that have perfect health. This latter point is particularly important. The idea is that if two individuals are both perfectly healthy, then preferences "should not matter" in determining the desirability of an income transfer. With two individuals at the same mediocre health level, it may be legitimate for the rich to claim that he is in a worse situation because he cares more for his health. This reasoning is not at all convincing, however, if he is in perfect health.

We now have the following result.

Proposition 4 *If the social ordering* $\mathbf{R}(.)$ *satisfies Weak Pareto, Independence, and Pigou-Dalton, then* $x_N \mathbf{P}(P_N^*) x_N'$ *whenever* $\min_i E^{\inf}(x_i, P_i^*) > \min_i E^{\sup}(x_i', P_i^*)$.

It is possible to refine the ordering by adding a safety axiom once again.

Super Safety: If there is i and $P_i^{*\prime} \supseteq P_i^*$ such that $x_N \mathbf{P}(P_i^{*\prime}, P_{N\setminus\{i\}}^{*\prime}) x_N'$ for all $P_{N\setminus\{i\}}^{*\prime} \supseteq P_{N\setminus\{i\}}^*$, and for no j and $P_j^{*\prime} \supseteq P_j^*$ one has $x_N' \mathbf{P}(P_j^{*\prime}, P_{N\setminus\{j\}}^{*\prime}) x_N$ for all $P_{N\setminus\{j\}}^{*\prime} \supseteq P_{N\setminus\{j\}}^*$, then $x_N \mathbf{P}(P_N^*) x_N'$.

This axiom is similar to the super safety principle of the previous section. It makes it possible to refine the ordering but not in a very simple way, because the logic of refinement is quite different in the social evaluation context, as compared to interpersonal comparisons. In interpersonal comparisons, one can refine one agent's preferences without refining the other agent's preferences, therefore only one term of the comparison is altered. In the social context, refining one agent's preferences alters the evaluation of the two allocations to be compared. We can make two observations.

First, the (complete) ordering defined by $x_N \mathbf{R}(P_N^*) x_N'$ iff $\min_i E^{\inf}(x_i, P_i^*) \geq \min_i E^{\inf}(x_i', P_i^*)$ satisfies the four axioms.

Second, even if $\min_i E^{\inf}(x_i, P_i^*) > \min_i E^{\inf}(x_i', P_i^*)$ and $\min_i E^{\sup}(x_i, P_i^*) > \min_i E^{\sup}(x_i', P_i^*)$, the Super Safety axiom, in conjunction with the other three, does not imply that $x_N \mathbf{P}(P_N^*) x'_N$ in general. To see this, consider a case in which there is one agent i who is far worse-off than the others, so that the evaluation depends only on his preferences. If $x_i \in NC(x'_i, P_i^*)$, it may happen nonetheless that $E^{\inf}(x_i, P_i^*) > E^{\inf}(x'_i, P_i^*)$ and $E^{\sup}(x_i, P_i^*) > E^{\sup}(x'_i, P_i^*)$. This is compatible with finding $P_i^{*\prime} \supseteq P_i^*$ such that $x_i P_i^{*\prime} x'_i$ and $P_i^{*\prime\prime} \supseteq P_i^*$ such that $x'_i P_i^{*\prime\prime} x_i$. Therefore the Super Safety axiom has no bite in this case.[15]

We however obtain an interesting refinement, as follows.

Proposition 5 *If the social ordering* $\mathbf{R}(.)$ *satisfies Weak Pareto, Independence, Pigou-Dalton, and Super Safety, then* $x_N \mathbf{P}(P_N^*) x'_N$ *whenever* $\min_i E^{\inf}(x_i, P_i^*) > \min_i E^{\inf}(x'_i, P_i^*)$ *and for every* j *such that* $E^{\inf}(x'_j, P_j^*) < \min_i E^{\sup}(x_i, P_i^*)$, $x_j P_j^* x'_j$.

Of course, the social rankings that have been derived are incomplete. Yet, they are finer than the Pareto-ranking proposed by Bernheim and Rangel (2009) – and they allow to introduce distributional considerations in welfare analysis, even if one only uses information about ordinal and non-complete preferences. While it certainly would be worthwhile to explore further the potential contribution of imposing additional ethical requirements, the path that could be taken is clearly traced out.

5 Conclusion

We have argued in this paper that it is possible to define a concept of interpersonally comparable well-being that uses only information about ordinal preferences – even if these

[15] This example cannot occur if one assumes that every preference P^* is the intersection of a set $\mathcal{B}(P^*)$ of strict preference relations which are the asymmetric parts of monotonic, transitive, and complete relations, and that every pair of preferences in $\mathcal{B}(P^*)$ satisfies the single-crossing property (i.e., the corresponding indifference curves cross at most once). But even under this domain restriction, the condition $\min_i E^{\inf}(x_i, P_i^*) > \min_i E^{\inf}(x'_i, P_i^*)$ and $\min_i E^{\sup}(x_i, P_i^*) > \min_i E^{\sup}(x'_i, P_i^*)$ cannot be sufficient to ensure $x_N \mathbf{P}(P_N^*) x'_N$. For instance, one may have two agents i, j who are far worse-off than the others, with $x_i \in NC(x'_i, P_i^*)$ and

$$\begin{aligned} E^{\inf}(x'_i, P_i^*) &< E^{\inf}(x_i, P_i^*) < E^{\inf}(x'_j, P_j^*) < E^{\sup}(x'_j, P_j^*) \\ &< E^{\sup}(x_i, P_i^*) < E^{\sup}(x'_i, P_i^*) < E^{\inf}(x_j, P_j^*) < E^{\sup}(x_j, P_j^*). \end{aligned}$$

One may in addition find $\bar{P}_i^* \supseteq P_i^*$ such that $E^{\sup}(x_i, \bar{P}_i^*) < E^{\inf}(x'_i, \bar{P}_i^*) < E^{\inf}(x'_j, P_j^*)$, thereby forcing to prefer x'_N no matter how one refines $P_{N\setminus\{i\}}^*$.

preferences are incomplete. Our paper therefore makes a contribution to two strands of the welfare economic literature. First, our introduction of incomplete preferences can be seen as an extension of the fair social choice approach. Second, we propose a method to define a normatively relevant concept of well-being as an extension of the Bernheim-Rangel (2009) approach to behavioral welfare economics. This makes it possible to go beyond Pareto-efficiency and introduce distributional considerations into the welfare evaluation. Of course, for our approach to be meaningful it is necessary to assume that individuals do have preferences over different features of life. However, it is not necessary that these preferences are complete, nor that the analyst has perfect information about them.

The interpersonal comparisons and social rankings we derive are unavoidably incomplete. Yet, if one refines the individual preferences, one reaches the standard approach with equivalent incomes as a limiting case. Moreover a more complete social ranking can also be obtained by imposing additional normative requirements. Further work should look for a definition of acceptable and feasible refinements – or for the development of better methods to measure preferences.

The well-being concept we propose is very different from traditional "subjective utility" or "happiness". We do not aim at measuring "true" happiness, but at formulating a concept that is meaningful for policy evaluation. Both the choice of the monotone path used in the equivalence approach and the choice of axioms to be imposed in the social evaluation exercise are essentially normative. This is not a weakness, but rather an advantage of the approach. When one aims at policy evaluation, it is better to make the underlying value judgments as open as possible. Having an informed debate about such value judgments in a formal model has always been the main objective of social choice theory.

References

Bernheim, D. 2009. Behavioral welfare economics. *Journal of the European Economic Association* 7, no. 2-3: 267-319.

Bernheim, D., and A. Rangel. 2009. Beyond revealed preference: choice theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics* 124, no. 1: 51-104.

Beshears, J., J. Choi, D. Laibson, and B. Madrian. 2008. How are preferences revealed? *Journal of Public Economics* 92, no. 8-9: 1787-94.

Blackorby, C., and D. Donaldson. 1988. Money metric utility: a harmless normalization? *Journal of Economic Theory* 46, no. 1: 120-129.

Blackorby, C., and D. Donaldson. 1990. A review article: the case against the use of the sum of compensating variations in cost-benefit analysis. *Canadian Journal of Economics* 23, no. 3: 471-94.

Brun, B., and B. Tungodden. 2004. Non-welfaristic theories of justice: Is the "intersection approach" a solution to the indexing impasse? *Social Choice and Welfare* 22, no. 1: 49-60.

d'Aspremont, C., and L. Gevers. 2002. Social welfare functionals and interpersonal comparability, in: K. Arrow, A. Sen, and K. Suzumura (eds.), *Handbook of Social Choice and Welfare, Vol. 1*, 459-541. Amsterdam: Elsevier.

Dalton, P. , and S. Ghosal. 2010. *Behavioral decisions and welfare.* CentER Tilburg: Discussion Paper 2010-143.

Deaton, A., and J. Muellbauer. 1980. *Economics and consumer behaviour.* Cambridge: Cambridge University Press.

Della Vigna, S. 2009. Psychology and economics: evidence from the field. *Journal of Economic Literature* 47: 315-72.

Dworkin, R. 1981. What is equality? Part 1: Equality of welfare. *Philosophy and*

Public Affairs 10: 185-246.

Fleurbaey, M. 2005. Health, wealth and fairness. *Journal of Public Economic Theory* 7, no. 2: 253-84.

Fleurbaey, M. 2009. Beyond GDP: the Quest for a Measure of Social Welfare. *Journal of Economic Literature* 47, no. 4: 1029-75.

Fleurbaey, M., and F. Maniquet. 2011. *A theory of fairness and social welfare.* Cambridge: Cambridge University Press.

Fleurbaey, M., E. Schokkaert, and K. Decancq. 2009. *What good is happiness?* CORE : Discussion Paper 2009/17.

Fleurbaey, M., and A. Trannoy. 2003. The impossibility of a Paretian egalitarian. *Social Choice and Welfare* 21, no. 2: 243-63.

Frey, B., and A. Stutzer. 2002. What can economists learn from happiness research? *Journal of Economic Literature* 40, no. 2: 402-35.

Gustafsson J. 2011, "An extended framework for preference relations", *Economics and Philosophy* 27:360-367.

Kahneman, D., and R. Sugden. 2005. Experienced utility as a standard of policy evaluation. *Environmental and Resource Economics* 32: 161-81.

Kahneman, D., P. Wakker, and R. Sarin. 1997. Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics* 112: 375-406.

King, M. 1983. Welfare analysis of tax reforms using household data. *Journal of Public Economics* 23: 183-214.

Köszegi, B., and M. Rabin. 2008. Choices, situations, and happiness. *Journal of Public Economics* 92: 1821-32.

Layard, R. 2005. *Happiness: lessons from a new science.* Allan Lane.

Loewenstein, G., and P. Ubel. 2008. Hedonic adaptation and the role of decision and experienced utility in public policy. *Journal of Public Economics* 92: 1795-810.

Noor, J. 2010. *Subjective welfare.* Department of Economics, Boston University.

Mimeo.

Rubinstein, A., and Y. Salant. 2012. Eliciting welfare preferences from behavioral datasets. Review of Economic Studies: forthcoming.

Salant, Y., and A. Rubinstein. 2008. (A,f): choice with frames. Review of Economic Studies 75: 1287-96.

Samuelson, P. 1974. Complementarity: an essay on the 40th anniversary of the Hicks-Allen revolution in demand theory. Journal of Economic Literature 12: 1255-1289.

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

——. 1985. *Commodities and capabilities*. Amsterdam: North-Holland.

Shafir, E., I. Simonson, and A. Tversky. 1993. Reason-based choice. *Cognition* 49: 11-36.

Sprumont, Y. 2011. Resource egalitarianism with a dash of efficiency. *Journal of Economic Theory* : forthcoming.

Tversky, A., and I. Simonson. 1993. Context-dependent preferences. *Management Science* 39, no. 10: 1179-89.

Willig, R. 1981. Social welfare dominance. *American Economic Review (Papers and Proceedings)* 71, no. 2: 200-204.

Appendix

Proof of Prop. 1 We first prove that B is a monotone path. As there is P^* such that $NC(0,P^*)=\{0\}$, and as B is such that $NC(0,P^*)\cap B\neq\varnothing$, necessarily $0\in B$.

Let $z,z'\in B$ be such that neither $z\geq z'$ nor $z\leq z'$. There is P^* such that zP^*z' and $P^{*\prime}$ such that $z'P^{*\prime}z$. By the preference principle, $(z,P^*)\succ(z',P^*)$ and $(z',P^{*\prime})\succ(z,P^{*\prime})$. By the restricted dominance principle, $(z,P^*)\sim(z,P^{*\prime})$ and $(z',P^*)\sim(z',P^{*\prime})$. This violates transitivity.

The fact that for all (x,P^*), $NC(x,P^*)\cap B\neq\varnothing$ then directly implies that B is unbounded and connected.

Let $(x,P^*),(x',P^{*\prime})$ be such that $LC(x,P^*)\cap UC(x',P^{*\prime})\cap B\neq\varnothing$. Let $z\in LC(x,P^*)\cap UC(x',P^{*\prime})\cap B$. By the preference principle, $(x,P^*)\succ(z,P^*)$ and $(z,P^{*\prime})\succ(x',P^{*\prime})$. By the restricted dominance principle, $(z,P^*)\sim(z,P^{*\prime})$. By transitivity, $(x,P^*)\succ(x',P^{*\prime})$. ■

Proof of Prop. 2 Let $(x,P^*),(x',P^{*\prime})$ be such that $LC(x,P^*)\cap(X\setminus LC(x',P^{*\prime}))\cap B\neq\varnothing$. As $LC(x,P^*)$ is open and lower comprehensive, while $X\setminus LC(x',P^{*\prime})$ is closed and upper comprehensive, $LC(x,P^*)\cap(X\setminus LC(x',P^{*\prime}))\cap B$ is not a singleton and there exist $z>z'$ in $LC(x,P^*)\cap(X\setminus LC(x',P^{*\prime}))\cap B$. As z is not on the lower boundary of $X\setminus LC(x',P^{*\prime})$, there is a refinement $\bar{P}^{*\prime}\supseteq P^{*\prime}$ such that $z\in UC\left(x',\bar{P}^{*\prime}\right)$. For all refinements $\bar{P}^*\supseteq P^*$, $x\bar{P}^*z$ because xP^*z. By Prop. 1, $z\in LC(x,\bar{P}^*)\cap UC(x',\bar{P}^{*\prime})\cap B$ implies that $(x,\bar{P}^*)\succ(x',\bar{P}^{*\prime})$. By the safety principle, $(x,P^*)\succ(x',P^{*\prime})$. ■

Proof of Prop. 3 There are two possible cases. First, let $(x,P^*),(x',P^{*\prime})$ be such that $LC(x',P^{*\prime})\cap(X\setminus LC(x,P^*))\cap B=\varnothing$, $UC(x,P^*)\cap(X\setminus UC(x',P^{*\prime}))\cap B=\varnothing$, and $LC(x,P^*)\cap(X\setminus LC(x',P^{*\prime}))\cap B\neq\varnothing$.

Let $z\in LC(x,P^*)\cap(X\setminus LC(x',P^{*\prime}))\cap B$ be such that z is not in the lower boundary of $X\setminus LC(x',P^{*\prime})$.

There is $\bar{P}^{*\prime}\supseteq P^{*\prime}$ such that for all $\bar{P}^*\supseteq P^*$, $z\in LC(x,\bar{P}^*)\cap UC(x',\bar{P}^{*\prime})\cap B$

(implying $(x,\bar{P}^*) \succ (x',\bar{P}^{*\prime})$ by Prop. 1). This implies that there is no $\bar{P}^* \supseteq P^*$ such that for all $\bar{P}^{*\prime} \supseteq P^{*\prime}$, $(x,\bar{P}^*) \prec (x',\bar{P}^{*\prime})$.

It remains to check that there exists no $\bar{P}^{*\prime} \supseteq P^{*\prime}$ such that for all $\bar{P}^* \supseteq P^*$, $(x,\bar{P}^*) \prec (x',\bar{P}^{*\prime})$. This directly follows from $UC(x,P^*) \cap (X \setminus UC(x',P^{*\prime})) \cap B = \varnothing$.

By the super safety principle, $(x,P^*) \succ (x',P^{*\prime})$.

The case in which $LC(x',P^{*\prime}) \cap (X \setminus LC(x,P^*)) \cap B = \varnothing$, $UC(x,P^*) \cap (X \setminus UC(x',P^{*\prime})) \cap B = \varnothing$, and $UC(x',P^{*\prime}) \cap (X \setminus UC(x,P^*)) \cap B \neq \varnothing$ is dealt with similarly. ■

Proof of Prop. 4 : Let x_N, x'_N be such that $\min_i E^{\inf}(x_i,P_i^*) > \min_i E^{\sup}(x'_i,P_i^*)$. Figure 7 illustrates the proof.

There exist $\hat{x}_N, \hat{x}'_N$ such that for all $i \in N$, $\hat{h}_i = \hat{h}'_i = 1$, $x_i P_i^* \hat{x}_i$, $\hat{x}'_i P_i^* x'_i$, and

$$\min_i E^{\inf}(x_i,P_i^*) > \min_i \hat{c}_i > \min_i \hat{c}'_i > \min_i E^{\sup}(x'_i,P_i^*).$$

Moreover, one can construct $\hat{x}_N, \hat{x}'_N$ so that there is a unique i_0 such that $\hat{c}_{i_0} = \min_i \hat{c}_i$ and $\hat{c}'_{i_0} = \min_i \hat{c}'_i$, and so that $\hat{x}'_i P_i^* \hat{x}_i$ for all $i \neq i_0$.

There exist $\bar{x}_{i_0}, \bar{x}'_{i_0}$ such that $\bar{h}_{i_0} = \bar{h}'_{i_0} < 1$ and $\hat{x}_{i_0} P_{i_0}^* \bar{x}_{i_0} P_{i_0}^* \bar{x}'_{i_0} P_{i_0}^* \hat{x}'_{i_0}$. For each $i \neq i_0$, let $\bar{x}_i, \bar{x}'_i$ be such that $\bar{h}_i = \bar{h}'_i = \bar{h}_{i_0}$, $\bar{x}'_i P_i^* \hat{x}'_i$ and $\bar{c}'_i - \bar{c}_i = \left(\bar{c}_{i_0} - \bar{c}'_{i_0}\right) / (n-1)$.

There exist $\bar{x}'''_{i_0}$ such that $\bar{h}'''_{i_0} = 1$ and $\hat{x}_{i_0} P_{i_0}^* \bar{x}'''_{i_0} P_{i_0}^* \bar{x}_{i_0}$. For each $i \neq i_0$, let $\bar{x}''_i, \bar{x}'''_i$ be such that $\bar{h}''_i = \bar{h}'''_i = 1$, $\hat{x}_i > \bar{x}'''_i > \bar{x}''_i > \hat{x}_{i_0}$ and $\bar{c}'''_i - \bar{c}''_i = \left(\hat{c}_{i_0} - \bar{c}'''_{i_0}\right) / (2(n-1))$. Let $\bar{x}''_{i_0} = \left(\hat{x}_{i_0} + \bar{x}'''_{i_0}\right)/2$. One has $\bar{c}'''_i - \bar{c}''_i = \left(\bar{c}''_{i_0} - \bar{c}'''_{i_0}\right) / (n-1)$.

Let $P_{i_0}^{*\prime} = P_{i_0}^*$ and for $i \neq i_0$, let $P_i^{*\prime}$ be such that $\bar{x}'_i P_i^{*\prime} \hat{x}'_i$, $\bar{x}'''_i P_i^{*\prime} \bar{x}_i$, $NC(\bar{x}'_i, P_i^{*\prime}) \cap NC(\hat{x}'_i, P_i^*) = \varnothing$, $NC(\hat{x}_i, P_i^*) \cap NC(\bar{x}''_i, P_i^{*\prime}) = \varnothing$, $NC(\bar{x}_i, P_i^{*\prime}) \cap NC\left(\bar{x}_{i_0}, P_{i_0}^{*\prime}\right) = \varnothing$.

For $i \neq i_0$, let $P_i^{*\prime\prime}$ be such that $NC(\bar{x}'_i, P_i^{*\prime\prime}) = NC(\bar{x}'_i, P_i^{*\prime})$, $NC(\bar{x}_i, P_i^{*\prime\prime}) = NC(\bar{x}_i, P_i^{*\prime})$, and for all x such that $\bar{x}'_{i_0} \leq x \leq \bar{x}_{i_0}$, $NC(x, P_i^{*\prime\prime}) = NC\left(x, P_{i_0}^{*\prime}\right)$.

Number the agents $i \neq i_0$ from 1 to $n-1$. By Pigou-Dalton,

$$\left(\bar{x}_1, \bar{x}'_2, ..., \bar{x}'_{n-1}, \bar{x}'_{i_0} + \bar{x}'_1 - \bar{x}_1\right) \mathbf{R}(P_1^{*\prime\prime}, P_2^{*\prime}, ..., P_{n-1}^{*\prime}, P_1^{*\prime\prime}) \bar{x}'_N,$$

and by independence,

$$\left(\bar{x}_1, \bar{x}'_2, ..., \bar{x}'_{n-1}, \bar{x}'_{i_0} + \bar{x}'_1 - \bar{x}_1\right) \mathbf{R}(P^{*\prime}_N)\bar{x}'_N.$$

Repeating this argument for agent 2, one obtains

$$\left(\bar{x}_1, \bar{x}_2, \bar{x}'_3, ..., \bar{x}'_{n-1}, \bar{x}'_{i_0} + 2\left(\bar{x}'_1 - \bar{x}_1\right)\right) \mathbf{R}(P^{*\prime}_N)\bar{x}'_N.$$

After applying this argument also to $i = 3, ..., n-1$, and noting that $(n-1)\left(\bar{x}'_1 - \bar{x}_1\right) = \bar{x}_{i_0} - \bar{x}'_{i_0}$, one obtains $\bar{x}_N\mathbf{R}(P^{*\prime}_N)\bar{x}'_N$.

By weak Pareto, $\bar{x}'''_N\mathbf{P}(P^{*\prime}_N)\bar{x}_N$. By $n-1$ applications of Pigou-Dalton, $\bar{x}''_N\mathbf{R}(P^{*\prime}_N)\bar{x}'''_N$. By transitivity, $\bar{x}''_N\mathbf{P}(P^{*\prime}_N)\bar{x}'_N$.

Let $P^{*\prime\prime\prime}_{i_0} = P^*_{i_0}$ and for $i \neq i_0$, let $P^{*\prime\prime\prime}_i$ be such that $NC\left(\bar{x}'_i, P^{*\prime\prime\prime}_i\right) = NC\left(\bar{x}'_i, P^{*\prime}_i\right)$, $NC\left(\bar{x}''_i, P^{*\prime\prime\prime}_i\right) = NC\left(\bar{x}''_i, P^{*\prime}_i\right)$, and $NC\left(\hat{x}_i, P^{*\prime\prime\prime}_i\right) = NC\left(\hat{x}_i, P^*_i\right)$, $NC\left(\hat{x}'_i, P^{*\prime\prime\prime}_i\right) = NC\left(\hat{x}'_i, P^*_i\right)$.

By independence, $\bar{x}''_N\mathbf{P}(P^{*\prime\prime\prime}_N)\bar{x}'_N$.

By weak Pareto, $\hat{x}_N\mathbf{P}(P^{*\prime\prime\prime}_N)\bar{x}''_N$ and $\bar{x}'_N\mathbf{P}(P^{*\prime\prime\prime}_N)\hat{x}'_N$. By transitivity, $\hat{x}_N\mathbf{P}(P^{*\prime\prime\prime}_N)\hat{x}'_N$.

By independence, $\hat{x}_N\mathbf{P}(P^*_N)\hat{x}'_N$. By weak Pareto, $x_N\mathbf{P}(P^*_N)\hat{x}_N$ and $\hat{x}'_N\mathbf{P}(P^*_N)x'_N$. By transitivity, $x_N\mathbf{P}(P^*_N)x'_N$.

■

Proof of Prop. 5 : Let x_N, x'_N be such that $\min_i E^{\inf}(x_i, P^*_i) > \min_i E^{\inf}(x'_i, P^*_i)$ and for every j such that $E^{\inf}(x_j, P^*_j) < \min_i E^{\sup}(x'_i, P^*_i)$, $x_j P^*_j x'_j$.

There is j_0 and a complete $\bar{P}^*_{j_0} \supseteq P^*_{j_0}$ such that $E(x'_{j_0}, \bar{P}^*_{j_0}) = \min_i E^{\inf}(x'_i, P^*_i)$. Take any $\bar{P}^*_i \supseteq P^*_i$ for all $i \neq j_0$. Necessarily, $\min_i E^{\sup}(x'_i, \bar{P}^*_i) = E(x'_{j_0}, \bar{P}^*_{j_0}) < \min_i E^{\inf}(x_i, P^*_i) \leq \min_i E^{\inf}(x_i, \bar{P}^*_i)$, implying that $x_N\mathbf{P}(\bar{P}^*_N)x'_N$ by Prop. 4.

Suppose there were k and $P^{*\prime}_k \supseteq P^*_k$ such that $x'_N\mathbf{P}(P^{*\prime}_k, P^{*\prime}_{N\setminus\{k\}})x_N$ for all $P^{*\prime}_{N\setminus\{k\}} \supseteq P^*_{N\setminus\{k\}}$. By the previous paragraph, it is impossible that $k \neq j_0$, because with $P^{*\prime}_{j_0} = \bar{P}^*_{j_0}$ one would then have $x_N\mathbf{P}(P^{*\prime}_k, P^{*\prime}_{N\setminus\{k\}})x'_N$. Therefore $k = j_0$. There is no loss in generality in assuming that all $P_i^{*\prime}$ $(i \in N)$ are complete[16] when one writes that $x'_N \mathbf{P}(P_k^{*\prime}, P_{N\setminus\{k\}}^{*\prime}) x_N$ for all $P_{N\setminus\{k\}}^{*\prime} \supseteq P_{N\setminus\{k\}}^{*}$. Necessarily $P_{j_0}^{*\prime} \neq \bar{P}_{j_0}^{*}$ and $E(x'_{j_0}, P_{j_0}^{*\prime}) \geq \min_i E^{\inf}(x_i, P_i^*)$, otherwise $x_N \mathbf{P}(P_k^{*\prime}, P_{N\setminus\{k\}}^{*\prime}) x'_N$ would be guaranteed.

For all j such that $E^{\inf}(x_j, P_j^*) < \min_i E^{\sup}(x'_i, P_i^*)$, let $P_j^{*\prime} \supseteq P_j^*$ be a (complete) ordering such that $E(x_j, P_j^{*\prime}) = E^{\inf}(x_j, P_j^*)$. This set of j may include j_0. Necessarily, $\min_i E^{\inf}(x_i, P_i^{*\prime}) = \min_i E^{\sup}(x_i, P_i^{*\prime})$, which can be denoted $E(x_j, P_j^{*\prime})$, for one of these j.

Moreover, for all of them, $E(x'_j, P_j^{*\prime}) = E^{\inf}(x'_j, P_j^{*\prime}) = E^{\sup}(x'_j, P_j^{*\prime}) \geq \min_i E^{\sup}(x'_i, P_i^{*\prime})$.

Now, for all of them, $x_j P_j^* x'_j$, which implies $x_j P_j^{\prime *} x'_j$ and therefore $E(x_j, P_j^{*\prime}) > E(x'_j, P_j^{*\prime})$, implying that $\min_i E^{\inf}(x_i, P_i^{*\prime}) > \min_i E^{\sup}(x'_i, P_i^{*\prime})$, and therefore $x_N \mathbf{P}(P_N^{*\prime}) x'_N$ by Prop. 4. One obtains a contradiction with the assumption that $x'_N \mathbf{P}(P_N^{*\prime}) x_N$.

Therefore super safety applies, and one concludes that $x_N \mathbf{P}(P_N^*) x'_N$. ■

[16] Completeness of P^* means that there is a complete binary relation R^* such that xR^*y iff not(yP^*x).

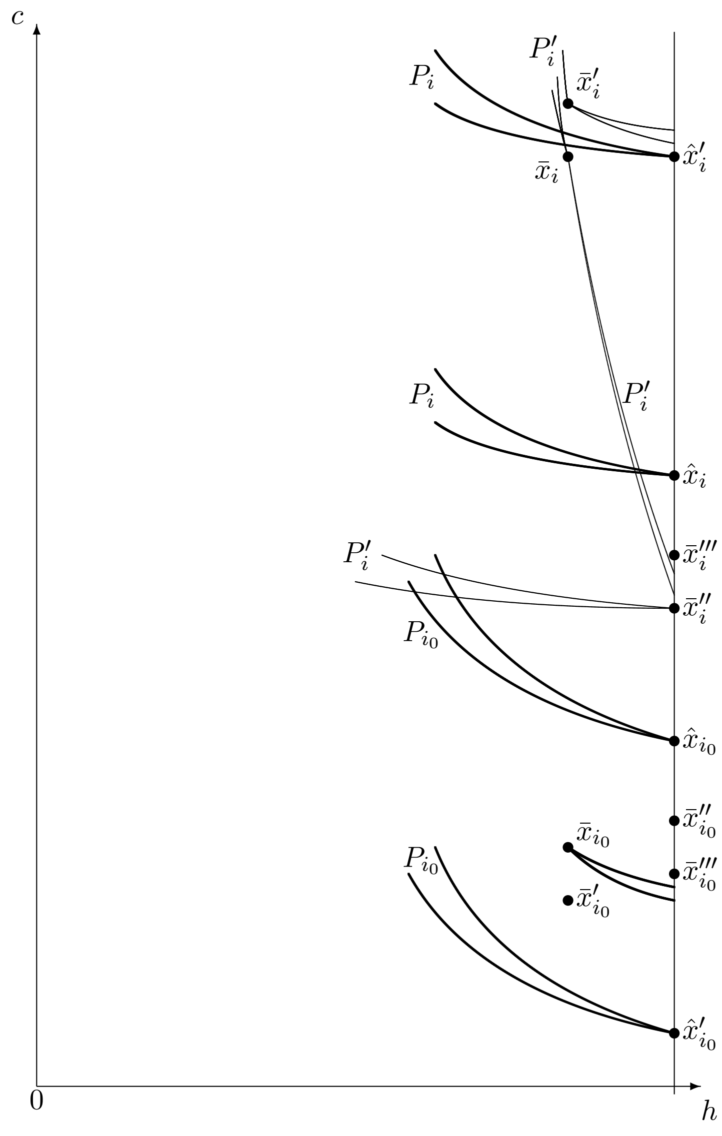

Figure 7: Illustration of the proof of Prop. 4.