HERMAN DELEECK
CENTRE FOR SOCIAL POLICY

Koen Decancq & Annemie Nys

Non-parametric well-being comparisons

WORKING PAPER

No. 18.14

September 2018

Universiteit Antwerpen

University of Antwerp
Herman Deleeck Centre for Social Policy
centrumvoorsociaalbeleid.be

Non-parametric well-being comparisons

Koen Decancq & Annemie Nys

Working Paper No. 18 / 14
September 2, 2018

ABSTRACT

We study the problem of making interpersonal well-being comparisons when individuals have heterogeneous – possibly incomplete – preferences. We present a robust – also incomplete – criterion for well-being comparisons that states that one individual is better off than another one if the intersection between the extended upper contour set of the better off individual and the extended lower contour set of the worse off individual is empty. We implement the criterion in the consumption-health space using an online survey with 2,260 respondents in the United States to investigate how incomplete the resulting interpersonal well-being comparison is. To chart the contour sets of the respondents, we propose a new "adaptive bisectional dichotomous choice" (ABDC) procedure that is based on a limited number of dichotomous choices and some mild non-parametric assumptions on the preferences. While the ABDC procedure does not reject that the preferences of a large majority of the respondents satisfy these non-parametric assumptions, it has sufficient power to reject several standard parametric assumptions such as linearity or Cobb-Douglas preferences for an overwhelming number of respondents. Finally, we find that about one fifth of all pairs of respondents can be ranked in a robust way with the proposed criterion. A more complete version of the criterion is able to rank more than half of the pairs.

Corresponding author:

Koen Decancq
Centre for Social Policy (CSB), University of Antwerp
Centre for Philosophy of Natural and Social Science, London School of Economics
Department of Economics, KULeuven
CORE, Université catholique de Louvain

1 Introduction

Interpersonal well-being comparisons constitute an indispensable component of the design and evaluation of most public policies. As real-world policies rarely lead to Pareto improvements, gains in well-being for some individuals need to be balanced against losses for others. This balancing requires a method to make interpersonal well-being comparisons.

Recently, a consensus has emerged in the literature that well-being is best seen as a multidimensional notion. In their report on the measurement of economic performance and social progress, Stiglitz et al. (2009) have argued the inclusion of non-monetary dimensions such as health, educational achievements, and employment status in an analysis of well-being. Yet, as soon as we move into a multidimensional framework, making interpersonal well-being comparisons becomes more complicated than when considering a single dimension in isolation. One individual may be better off in one dimension whereas another individual may be better off in another. Which individual should then be considered better off overall?

To make multidimensional well-being comparisons, a procedure is therefore needed to aggregate the outcomes in the different dimensions. Such a procedure can be based on an objective composite well-being index, subjective measures of well-being, or the preferences of the concerned individuals themselves (see Fleurbaey and Blanchet (2013) and Decancq et al. (2015b) for recent surveys). In line with the mainstream approach in economics, we start from the premise that interpersonal well-being comparisons should respect individual preferences, acknowledging that these preferences are not necessarily complete, nor uniform across people.[1]

We propose to use the *Nested Contours* criterion to make interpersonal well-being comparisons. This criterion has its roots in the theory of fair allocation (see Fleurbaey and Maniquet (2011, 2017)). According to the criterion, one individual is considered better off than another one if the intersection between the extended upper contour set of the better-off individual and the extended lower contour set of the worse-off individual is empty. The extended upper contour set of an individual contains all life situations that she prefers over her current

[1] This premise clearly entails a value judgement on the nature of well-being. Since the early work of Robbins (1932 [1984], 1938) it is well-known that value judgements are inevitable when making interpersonal well-being comparisons (see Hammond (1991) for a survey).

life situation as well as all life situations that she cannot compare to her current life situation. The lower contour set is the complement of the extended upper contour set and vice versa. When both individuals have complete preferences, the *Nested Contours* criterion coincides with the natural idea that the individual whose indifference curve lies everywhere above the indifference curve of another individual, is better off. Because the indifference curves of two individuals can cross, the *Nested Contours* criterion is incomplete. It is therefore unclear how useful the criterion is to make real-world interpersonal well-being comparisons. In this paper we are curious what the chances are that two randomly selected individuals can be ranked by means of the *Nested Contours* criterion.

To compare the well-being of real-world individuals with the *Nested Contours* criterion, we need an operational method to draw indifference curves or contours sets for each individual. The standard parametric method to chart indifference curves would be to estimate the parameters of an *a priori* chosen parametric utility function using observational or hypothetical data. Such a parametric approach is vulnerable to misspecification of the utility function and does generally not allow to draw individual-specific indifference curves. We rely therefore on an alternative –entirely non-parametric– method to chart individual indifference curves. This method is inspired by the early experimental work by Thurstone (1931) and MacCrimmon and Toda (1969) on the shape of individual indifference curves.[2] It introduces techniques from revealed preference demand analysis, developed in the wake of the work by Samuelson (1948) and Varian (1982), in a setting with non-market goods such as health. We call it the "Adaptive Bisectional Dichotomous Choice" (ABDC) method.

The ABDC method consists of two steps. In the first step, respondents are presented a series of dichotomous choices between pairs of life situations that consist of their actual life situation and a hypothetical one. The hypothetical life situations are obtained by the so-called adaptive bisectional algorithm. The adaptive bisectional algorithm proceeds iteratively and generates each time a hypothetical life situation which is situated in the middle of the interval where the indifference curve should be, based on the responses in previous choices.[3] In the second step, the choices of the respondent and the non-parametric assump-

[2] Moscati (2007) provides an historical overview of the experimental literature to chart individual indifference curves.

[3] The adaptive bisectional algorithm has been used in a parametric framework to elicit individual risk and time preference parameters (Hardisty et al., 2013; Toubia et al., 2013), to elicit preferences for health states in time trade-off or standard gamble methods (Prades et al., 2014), or in "double-bounded" discrete choice experiments (Watson and Ryan, 2007).

tions of monotonicity and –when desired– convexity of the preferences provide bounds around the indifference curves of the respondents.[4] The main novelty of the ABDC method is the combination of both steps. This allows us to obtain bounds on individual indifference curves in an entirely non-parametric way in a non-market setting, based on a limited number of dichotomous choices. The latter feature permits a large-scale implementation of the method in a standard online survey.

We illustrate the implementation of the ABDC method with data from an online survey that has been carried out in January 2017 with 2,260 respondents from the US. We focus on presumably two of the most important dimensions of well-being: consumption and health. The ABDC method allows us to provide bounds on the indifference curves and to test several parametric and non-parametric assumptions about the preferences of the respondents. While the ABDC procedure does not reject that the preferences of a large majority of the respondents satisfy our non-parametric assumptions of monotonicity and convexity, it has sufficient power to reject several standard parametric assumptions such as linearity or Cobb-Douglas preferences for an overwhelming number of respondents. We find that about one fifth of all pairs of respondents in our data set can be ranked in a robust way with the *Nested Contours* criterion. A more complete version of the criterion is able to rank more than half of all pairs.

The contribution of the paper is fourfold. First, we discuss the *Nested Contours* criterion for interpersonal well-being comparisons, and provide a new generalization of it. Second, we propose the ABDC method to chart individual indifference curves without having to rely on any parametric assumption about the underlying preferences. Third, we implement the ABDC method in a large-scale online survey and document the consistency between the responses and commonly made parametric and non-parametric assumptions about preferences in the consumption-health space. Finally, we measure the power and usefulness of *Nested Contours* to make interpersonal well-being comparisons and discuss which socio-demographic groups are considered better off according to this criterion.

The paper is structured as follows. Section 2 presents the framework and discusses our assumptions on the individual preferences in detail. The third section

[4] In a recent paper, Le Lec and Krawczyk (2018) elicit distributional preferences with a similar non-parametric procedure, using a large number of dichotomous choices in a so-called "multiple price list" format, see Andersen et al. (2006).

deals with the interpersonal well-being comparisons based on the *Nested Contours* criterion and its generalization. Section 4 discusses the data and the ABDC method to chart the indifference curves of the respondents. The results are presented in Section 5. First, we discuss the consistency between the responses and standard assumptions about preferences. Second, we present the results about the power of *Nested Contours* to make real-world interpersonal well-being comparisons. Section 6 concludes and sketches avenues for further research.

2 Individual preferences

The life situation of individual i is described by the m-dimensional vector ℓ_i. This vector captures her outcomes in the m relevant dimensions of life. For convenience, we assume that the set of life situations $\mathcal{L}$ is non-negative and bounded by the m-dimensional vector $\overline{\ell}$, i.e., $\ell_i \in \mathcal{L} = \left\{\ell_i \in \mathbb{R}^m_+ | \ell_i \leq \overline{\ell}\right\}$.[5]

Each individual has a binary relation P_i on the set of life situations that we call her preference relation. This preference relation captures her value judgements about which life situation she would prefer. In our setting, a preference relation is not necessarily revealed in actual choice behaviour, because not all dimensions of well-being are under the control of the individuals, such as their health status, for instance. Clearly, preferences may be different across people and can change over time. Below, we discuss four weak properties of the preference relation.

First, we assume that the preference relation is an asymmetric and transitive binary relation or a strict partial ordering.

Asymmetric. For all $\ell_i, \ell'_i \in \mathcal{L}$, if $\ell_i P_i \ell'_i$, then not $\ell'_i P_i \ell_i$.

Transitive. For all $\ell_i, \ell'_i, \ell''_i \in \mathcal{L}$, if $\ell_i P_i \ell'_i$ and $\ell'_i P_i \ell''_i$, then $\ell_i P_i \ell''_i$.

We call $\mathcal{P}$ the set of asymmetric and transitive binary relations. Although there is ample evidence that real-world preference relations are not always transitive (May, 1954; Tversky, 1969; Loomes et al., 1991), requiring that preference relations to be used in normative analysis are transitive seems a reasonable consistency requirement. Transitivity will play an important role in the rest of the paper.

[5] Let $<, \leq$, and $\ll$ denote the standard vector inequalities.

Note that we do not require the individual preference relations to be complete. Preferences can be incomplete for several reasons. First, preferences over alternative life situations are hypothetical by nature and may be unfamiliar and complex, so that the concerned individuals may feel that they cannot confidently compare all pairs of life situations (Butler and Loomes (1988, 2007) discuss so-called "imprecise preferences").[6] Alternatively, there may be a situation in which individual preferences are complete, but where the observer only manages to obtain incomplete information about these preferences. In this paper, we remain agnostic about the reason why individual preferences are incomplete in the eyes of the observer, but choose not to impose completeness from the outset.

Third, we assume the preference relation P_i to be monotone. This assumption imposes that the dimensions are "goods" and that individuals are not satiated, so that more is better. In other words, we assume that all individuals prefer a life situation with strictly more in at least one dimension, and not less in all other dimensions. This seems a reasonable assumption in the consumption-health space considered in this paper, provided that its bound $\overline{\ell}$ is not "too high".

Monotone. For all $\ell_i, \ell_i' \in \mathcal{L}$, if $\ell_i > \ell_i'$, then $\ell_i P_i \ell_i'$.

In addition, the preference relation P_i can be assumed to be convex. Convexity reflects a preference for well-balanced lives and is therefore more controversial. It requires furthermore that the dimensions of life are measured on a cardinal scale, which may not be appropriate in all contexts. Contrary to the other three properties, we will not assume that the preference relation is convex, but rather discuss its implications on our results.

Convex. For all $\ell_i, \ell_i', \ell_i'' \in \mathcal{L}$ and for all $0 < \alpha < 1$, if $\ell_i' P_i \ell_i$ and $\ell_i'' P_i \ell_i$, then $(\alpha \ell_i' + (1-\alpha)\, \ell_i'')\, P_i \ell_i$.

Finally, we define the upper and lower contour set of the preference relation P_i

[6] In a framework where choices between alternatives may depend on so-called frames, i.e., features of the choice environment rather than of the alternatives considered, Salant and Rubinstein (2008, p. 1292) define an asymmetric and transitive preference relation based on the choices that are consistent across all considered frames. Bernheim and Rangel (2009, p. 60) define a similar incomplete preference relation that is asymmetric and acyclic. For a discussion, see also Fleurbaey and Schokkaert (2013).

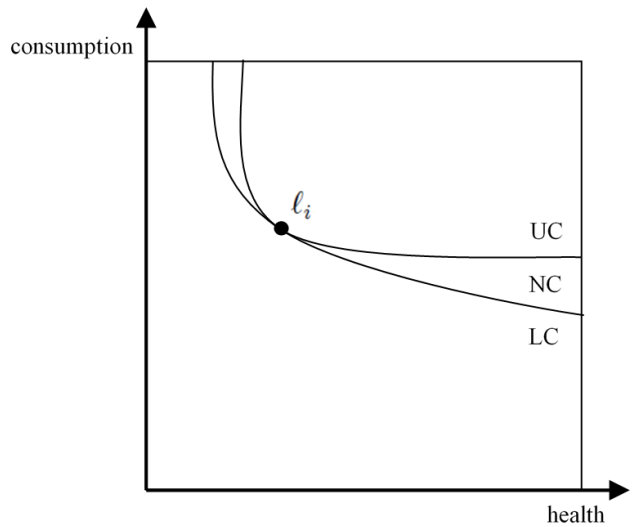

Figure 1: Upper and lower contour set and non-comparable set of individual i

in life situation ℓ_i as follows:

$$\begin{aligned} UC(\ell_i, P_i) &= \{\ell_i' \in \mathcal{L} | \ell_i' P_i \ell_i\} \\ LC(\ell_i, P_i) &= \{\ell_i' \in \mathcal{L} | \ell_i P_i \ell_i'\}. \end{aligned}$$

When the preference relation is convex, the upper contour set is convex as well. We define the non-comparable set $NC(\ell_i, P_i)$ as the set of all life situations that belong neither to $UC(\ell_i, P_i)$ nor to $LC(\ell_i, P_i)$

$$NC(\ell_i, P_i) = \mathcal{L} \setminus (UC(\ell_i, P_i) \cup LC(\ell_i, P_i)). \tag{1}$$

The more incomplete the preference relation P_i is, the larger the non-comparable set $NC(\ell_i, P_i)$. Monotonicity, however, imposes limitations on the size of the non-comparable set. See Figure 1 for an illustration. The more complete the individual preference relation is, the smaller the non-comparable set becomes. Based on the definitions of the contours sets, we define the extended upper and lower contour set as the union of the respective contour set and the non-

comparable set:

$$\begin{aligned} UC^{+}(\ell_i, P_i) &= UC(\ell_i, P_i) \cup NC(\ell_i, P_i) \\ LC^{+}(\ell_i, P_i) &= LC(\ell_i, P_i) \cup NC(\ell_i, P_i). \end{aligned} \tag{2}$$

Alternatively, the extended upper contour set is the complement of the lower contour set and vice versa.

3 Interpersonal well-being comparisons

We now address the question how to make interpersonal well-being comparisons in a way that is respectful of the individual preference relations.[7] Clearly, we cannot confine ourselves to comparing information about life situations alone. We need to include additional information on the preference relations of the individuals at hand and, hence, we will compare pairs that consist of a life situation and a preference relation for each individual (Fleurbaey and Blanchet, 2013). When the individual with the pair (ℓ_i, P_i) is considered better off than the individual with the pair (ℓ_j, P_j), we write $(\ell_i, P_i) \succ (\ell_j, P_j)$. We use the following criterion to make interpersonal comparisons:

Nested Contours. For all $\ell_i, \ell_j \in \mathcal{L}$ and $P_i, P_j \in \mathcal{P}$, if $UC^{+}(\ell_i, P_i) \cap LC^{+}(\ell_j, P_j) = \varnothing$, then $(\ell_i, P_i) \succ (\ell_j, P_j)$.

The *Nested Contours* criterion requires the extended upper contour set of the better off individual to be non-intersecting with the extended lower contour set of the worse off individual.[8] Clearly, the criterion leads to incomplete interpersonal comparisons: some individuals can be ranked, others cannot. Figure 2 provides an illustration. In the left-hand panel, the condition of the *Nested Contours* criterion is fulfilled, so that individual i (in black) is better off than

[7] In fact, the term "interpreference" well-being comparisons would be more appropriate. The proposed criterion can also be applied to intrapersonal well-being comparisons for an individual with an unstable (possibly incomplete) preference relation, see Weisbrod (1977). A preference relation may be unstable because it depends on the current life situation, for instance. Ubel et al. (2005) and Dolan and Kahneman (2008) provide empirical examples of persons who have different preference relations in different health states, and who fail moreover to anticipate their preference relation in other health states.

[8] The *Nested Contours* criterion is an extension of its namesake which has been introduced by Decancq et al. (2014) in a model with complete preferences. Fleurbaey and Maniquet (2017) study the connections of that criterion with the social choice literature on fair allocations and present various strengthenings.

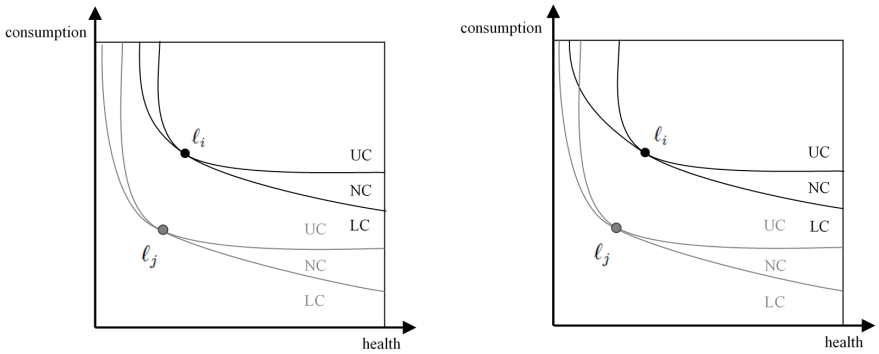

Figure 2: Nested contours at work. Nested contours implies a well-being ranking in the left-hand panel (of the black individual i over the grey individual j), but not in the right-hand panel

individual j (in grey). In the right-hand panel, on the contrary, the intersection between the extended upper and lower contours of the individuals is non-empty. In the non-empty intersection there are hypothetical life situations that individual i may prefer over her actual life situation. At the same time, individual j may prefer her actual life situation over the same hypothetical life situations. As a consequence, the hypothetical life situations are situated "between" individuals j and i. Likewise, there are other hypothetical life situations situated between both individuals with both individuals agreeing that the situation of individual i (in black) is preferred over the hypothetical situation, which is preferred over the situation of individual j. Hence, no robust and unambiguous well-being ranking can be derived based on *Nested Contours* alone.

Robbins (1932 [1984], 1938) and the literature in his wake have made clear that interpersonal comparisons are inevitably value-laden. It is therefore important to be precise and transparent about the value judgements that are embedded in the *Nested Contours* criterion. In line with the standard economic approach, only ordinal information about preferences is used. The criterion rules out any approach where well-being comparisons depend on a cardinal utility function or any other measure of the intensity of pleasure or preference satisfaction.[9] Well-

[9]Notably, the influential discussion by Robbins (1932 [1984], 1938) of interpersonal comparisons was centred on the nature of comparisons of the intensity of preference satisfaction across people. Famously, Robbins (1938, p. 637) quotes Jevons: "I see no means whereby such comparison can be accomplished. Every mind is inscrutable to every other mind and no common denominator of feeling is possible". Nevertheless, Robbins himself was convinced that

being comparisons based on nested contours are therefore not sensitive to the extent to which individuals are efficient in generating well-being, which avoids that individuals are considered as worse off because they have expensive tastes (see Arrow (1973); Sen (1985) amongst others).

As can be seen from the left-hand panel of Figure 2, the *Nested Contours* criterion requires that the non-comparable set $NC(\ell_i, P_i)$ of individual i does not intersect with the non-comparable set $NC(\ell_j, P_j)$ of individual j to make interpersonal well-being comparisons between both individuals. In fact, only the shapes of the non-comparable sets or, equivalently, the contour sets at the actual life situations matter for the well-being comparisons, and not the entire preference relation. This convenient feature of the criterion will turn out to be essential in the empirical part of this paper. Yet, as the entire shape of the the non-comparable set matters, considerably more information is used compared to approaches based on the axiom of Independence of Irrelevant Alternatives (see Arrow (1950) and Hansson (1973); Pazner (1979)).

The right-hand panel of Figure 2 shows that *Nested Contours* offers an incomplete criterion for interpersonal well-being comparisons and can be seen as an ethically robust, but minimal criterion for these comparisons. To the best of our knowledge, all preference-based well-being measures studied in the literature are consistent with the criterion.[10] Given its incompleteness, it is an open question how useful the criterion is to rank individuals in the real-world. To address this question, we propose a generalization of *Nested Contours* that completes it in a flexible way.

E_λ-Restricted Nested Contours For all $\ell_i, \ell_j \in \mathcal{L}$, and $P_i, P_j \in \mathcal{P}$, if $UC^+(\ell_i, P_i) \cap LC^+(\ell_j, P_j) \cap E_\lambda = \varnothing$, then $(\ell_i, P_i) \succ (\ell_j, P_j)$.

E_λ-Restricted Nested Contours builds on the same intuition as the original *Nested Contours* criterion, but requires the non-emptiness of the intersection between the extended upper and lower contour set and the evaluation set E_λ. One reason to restrict the criterion to an evaluation set, is when the policy

interpersonal well-being comparisons could not be avoided and that they should be explored "with the aid of the economist's technique", stressing that their "justification is more ethical than scientific".

[10] Prominent examples are the so-called ray utility measure (Samuelson, 1977; Pazner, 1979; Fleurbaey and Tadenuma, 2014), the distance function (Deaton, 1979), the money metric utility (McKenzie, 1957; Samuelson, 1974; King, 1983; Bosmans et al., 2018), or the equivalent income (Fleurbaey and Schokkaert, 2009; Decancq et al., 2015a,b).

maker considers hypothetical life situations outside the evaluation set as irrelevant from a normative perspective. Another reason to restrict the criterion is to use the size of the evaluation set as a measure of the robustness of the well-being comparison of two respondents. It is for this reason that we are introducing the restricted version of the criterion here.[11]

To be more precise, let us consider a sequence of nested evaluation sets $\mathcal{E} = (E_\lambda)$. Let each evaluation set in the sequence be a subset of the set of life situations $E_\lambda \subseteq \mathcal{L}$ and be indexed such that $E_\lambda \subset E_{\lambda'}$ whenever $\lambda < \lambda'$. The parameter λ provides a measure of the size of the evaluation set. The larger the evaluation set E_λ, the more demanding the E_λ-*Restricted Nested Contours* criterion is, and the more robust the well-being comparison becomes. Clearly, when λ is maximal and E_λ coincides with $\mathcal{L}$, the criterion is equivalent to *Nested Contours*. The largest λ for which E_λ-*Restricted Nested Contours* is fulfilled, quantifies the level of robustness of the interpersonal well-being criterion between individuals i and j. In the empirical part of this paper, we will study how large the maximal evaluation set in a sequence $\mathcal{E}$ is for a given pair of real-world individuals to get an idea of the robustness of their well-being ranking.

4 Charting contour sets

We now discuss how the E_λ-*Restricted Nested Contours* criterion can be implemented with real-world data. Although the criterion is applicable more generally, we restrict our attention to the case when $m = 2$, i.e., when a life situation can be described in two dimensions. This restriction helps to keep the problem tractable. We focus on the specific case when the life situation of individual i can be described by her consumption level c_i and her health status h_i, so that $\ell_i = (c_i, h_i)$. We describe first how we gathered information on the consumption level and health status. Then we introduce a method to chart the contour sets necessary to implement the criterion: the Adaptive Bisectional Dichotomous Choice (ABDC) method.

[11] Alternative measures of the robustness of the well-being comparison can be derived based on the size of the intersection between the extended upper and lower contour set.

4.1 Data

The data have been collected in the United States between the 17th and 24th of January 2017 by means of an on-line survey administered by the survey agency Qualtrics. The raw sample consists of 2,575 respondents between 25 and 85 years old. Given the sampling procedure and the on-line implementation of the survey, the sample is unlikely to be representative for the entire population of the United States. Indeed, white Americans are over-represented in the sample compared to census information, as well as people aged between 55 and 65 years. Appendix 1 presents some summary statistics.

To measure the individual consumption level of the respondents, we proceed as follows.[12] To each respondent we ask to report the monthly amount spent on various consumption items at the personal and household level, see Table 5 in Appendix 2. Moreover, for the household-level consumption items "food" and "transportation", respondents are asked to assign to each household member the share of the reported amount. The remaining household-level consumption items are assumed to be shared equally among all household members. To obtain an estimate of the individual consumption level, we sum the personal consumption items and the share of the household consumption items as indicated in Table 5. We ask respondents whether the computed amount seems reasonable to them as their individual consumption level on a five-point scale. About 4% of all respondents signal that the computed consumption level is "not at all" reasonable to them. These respondents have been removed from the sample. As discussed before, the set of life situations is assumed to be bounded by $\overline{\ell} = \left(\overline{c}, \overline{h}\right)$. We have set the upper bound for the consumption level at \$3,000, which approximately corresponds to the 90th percentile value of the consumption distribution. We remove the respondents from the sample whose consumption level is larger than the upper bound. The left-hand panel of Figure 3 presents the consumption distribution of the remaining sample of 2,260 respondents.

Health status is measured by a standard battery of health questions, the so-called 12-Item Short Form Health Survey (SF-12), see Ware et al. (1996). Table 6 in Appendix 3 provides more details. To convert the answers to these 12 questions into a single health index with which respondents can identify, we

[12] The questions used to measure individual consumption are very similar to ones in the ad-hoc module of the "Longitudinal Internet Studies for the Social sciences" (LISS), an on-line panel study carried out in the Netherlands. Cherchye et al. (2012) provide more information.

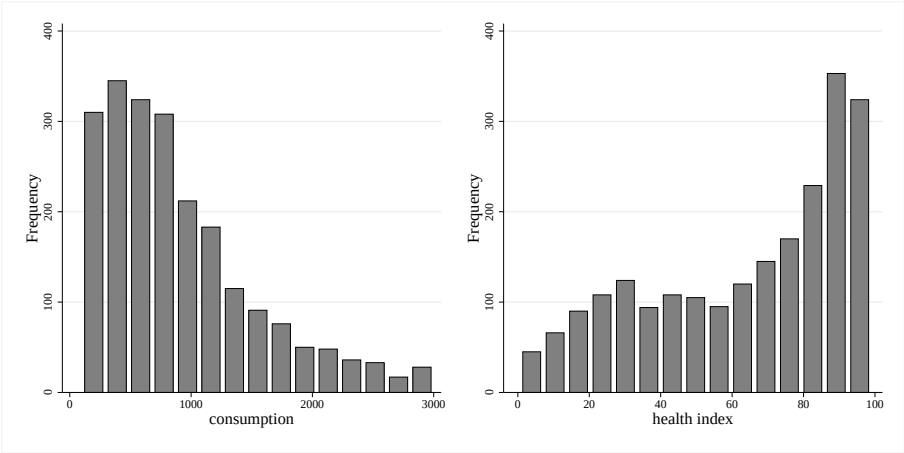

Figure 3: Distribution of individual consumption (left) and health index (right)

follow the following procedure. In a first step we convert the responses into two sub dimensions: emotional health and physical health. Each sub dimension is measured on a scale between 0 and 100 and is based on the responses to a set of five from the 12 questions. Then, in a second step, respondents are shown their personal scores on both sub dimensions and asked to divide 100 points over both sub dimensions reflecting their personal opinion on which sub dimension is relatively more important when considering their overall health status. About 37% of the respondents give equal weights to both sub dimensions and about 33% give more weight to the sub dimension physical health. Finally, an overall health index is computed as the weighted average of the scores on both sub dimensions. This health index is naturally bounded by its maximal value $\overline{h}$ of 100. We treat the obtained health indices as interpersonally comparable.[13] The distribution of the health index is presented in the right-hand panel of Figure 3.

4.2 Adaptive Bisectional Dichotomous Choice

We now discuss a method to determine the shape of the extended upper and lower contour set for each individual. We call it the Adaptive Bisectional Di-

[13]This assumption is a short-cut to reduce the dimensionality of the description of the life situations. Health status in itself can be argued to be a multidimensional notion. Comparing health status across individuals then raises similar questions as comparing well-being. In principle, the methods presented in this paper could be used to address that problem as well.

chotomous Choice (ABDC) method. The method consists of two steps. First, each respondent is confronted with several dichotomous choices, each consisting of her actual life situation ℓ_i and a hypothetical life situation ℓ_i'. Second, the extended upper and lower contour set for each individual are charted based on her choices and our assumptions on the preference relation.

The dichotomous choices in the first step of the method are organized in four sets. In each set of dichotomous choices, one dimension of the hypothetical life situation is fixed at a reference level. The so-called bisectional algorithm is applied on the other dimension.[14] In each iteration of the bisectional algorithm, the middle point of the interval is selected to which the non-comparable set belongs according to the answers in the previous iterations. Each time, the respondent is asked to compare this hypothetical life situation to her current life situation. This response narrows down the interval to which the non-comparable set belongs for the next iteration. More precisely, the bisectional algorithm proceeds iteratively and multiplies the level of the non-fixed dimension in the hypothetical life in each iteration t by a ratio r_t. This ratio is based on the decision of the respondent in the previous iteration, d_{t-1}, in the following way:

$$r_t = r_{t-1} + \frac{d_{t-1}}{2^t}, \tag{3}$$

where

$$\begin{cases} d_{t-1} = 1 & \text{if } \ell_i P_i \ell_i' \text{ in interation } t-1 \\ d_{t-1} = -1 & \text{otherwise,} \end{cases}$$

and $r_0 = 0$ and $d_0 = 1$. In the first iteration $(t = 1)$, we have that $r_t = 1/2$, and in the second iteration $(t = 2)$ we have that $r_t = 3/4$ when the respondent has chosen her own situation in the first iteration and $r_t = 1/4$ otherwise, and so on.

The left-hand panel of Figure 4 illustrates the bisectional algorithm of the ABDC method. The actual life situation of the respondent is indicated by $\ell_i = (c_i, h_i)$. In the first dichotomous choice the respondent is asked to compare her actual life situation with the hypothetical life situation that consists of the maximal

[14] Prades et al. (2014) compare the bisectional algorithm to alternatives, such as the titration and ping-pong algorithm. The titration algorithm starts from one end of the relevant interval and moves up (or down) until the respondent changes her preference for one option over the other, whereas in the ping-pong algorithm the respondent starts from one end of the relevant interval and moves to the other end, narrowing down the interval where indifference is expected to be reached. The authors conclude that the bisection algorithm usually performs better because it presents fewer hypothetical situations "far away" from the indifference curve.

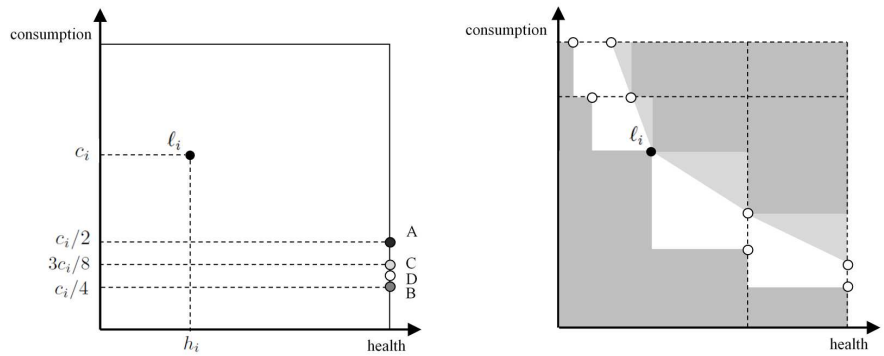

Figure 4: The ABDC method at work. The left-hand panel shows four hypothetical situations generated by the bisectional algorithm. The right-hand panel shows how the responses of the ABDC method lead to bounds on the non-comparable set of individual i

health level and a consumption level of $c_i/2$. This hypothetical life situation is indicated by the dark grey circle labelled A in the left-hand panel of Figure 4. Imagine that the respondent indicates that she prefers life situation A over her current life situation, then we know that life situation A belongs to her upper contour set and that the non-comparable set must be situated somewhere below life situation A. In the next dichotomous choice, the bisectional algorithm presents therefore a new hypothetical life situation with a lower consumption level, situated halfway between 0 and $c_i/2$, i.e., at $c_i/4$, as indicated in the figure by the life situation labelled B. Imagine that the respondent prefers her current situation to life situation B, then this life situation belongs to her lower contour set and the non-comparable set must be situated somewhere between the hypothetical life situations A and B. The third hypothetical life situation is then selected by the bisectional algorithm halfway between these two life situations, at a consumption level $3c_i/8$. Imagine that the respondent prefers this new life situation C over her actual life situation, then this life situation belongs to her upper contour set and the non-comparable set must be situated somewhere between life situation B and C. The fourth life situation, labelled D is therefore selected halfway between these two hypothetical life situations (at the consumption level $5c_i/16$). Imagine that the respondent answers "I don't know" in the comparison between her current life situation and life situation D, then the tightest bounds that we obtain on her non-comparable set are life

situations C (that belongs to her upper contour set) and life situation B (in her lower contour set).

As can be seen from Table 1, the four sets of dichotomous choices differ in the precise reference level that is chosen for its fixed dimension. This reference level is equal to the problem-free health status $\overline{h}$ in the first set, a health status halfway between the current health status and problem-free health status in the second set, maximal consumption $\overline{c}$ in the third set, and a consumption level halfway between the current consumption level and maximal consumption in the final set. In the first two sets of dichotomous choices, a sequence of four dichotomous choices is presented to the respondent, whereas in the last two sets only three dichotomous choices are presented, to avoid fatigue.[15]

Table 1: Hypothetical life situations

	set of dichotomous choices	consumption level	health level	number of choices
1	full health	$c_i \times r_t$	$\overline{h}$	4
2	intermediate health	$c_i \times r_t$	$h_i + (\overline{h} - h_i)/2$	4
3	full consumption	$\overline{c}$	$h_i \times r_t$	3
4	intermediate cons.	$c_i + (\overline{c} - c_i)/2$	$h_i \times r_t$	3

Note: the ratio r_t is determined iteratively according to the bisectional algorithm, see equation (3).

In each of these dichotomous choices, the respondent is asked to choose either her current life situation ℓ_i or the presented hypothetical life situation ℓ_i'. If she cannot make an unambiguous choice between both situations in the pair $\{\ell_i, \ell_i'\}$, the respondent can use a third response option. This third option is presented between both life situations on the screen and is labelled "I don't know".[16] About 45% of the respondents in the sample never use this option, whereas 5% uses it for all dichotomous choices. Figure 5 shows the percentage of respondents who use the "I don't know" response in each of the dichotomous choices, grouped together in the four sets of dichotomous choices by means of grey scales. We see that within each set of dichotomous choices the number of "I don't know"s increases, whereas it mildly decreases across all dichotomous choices. As the bisectional algorithm gradually confronts the respondents with

[15] To make the health index used in the hypothetical life situation comprehensible for the respondents, we provide under each dichotomous choice a so-called vignette. Each vignette describes the health outcomes corresponding to the hypothetical health index by means of the same questions of the SF-12 that have also been used to measure the actual health index of the respondent (see Table 6). More information is provided in Appendix 4.

[16] Figure 9 in Appendix 4 presents a screenshot and the precise formulation of the question.

hypothetical choices closer to their non-comparable sets, it is intuitive that the number of "I don't know"s increases within each set. Learning about their own preferences or the procedure to compare hypothetical choices may explain the overall decreasing share of "I don't know"s.

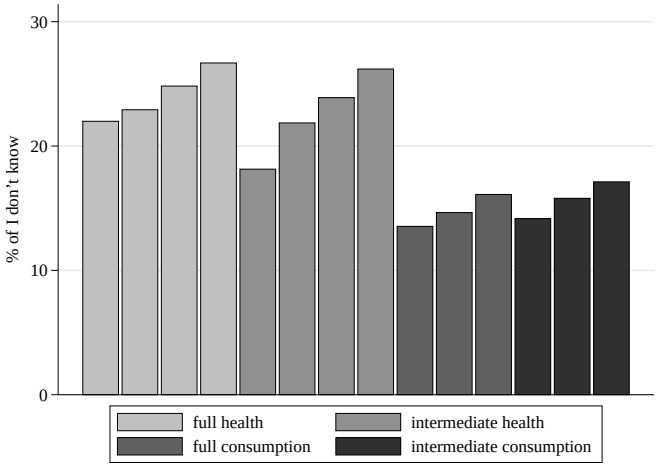

Figure 5: Percentage of "I don't know" responses for each dichotomous choice

In the second step of the ABDC method, the upper and lower contour set are charted based on the choices from the first step and our assumptions about the preference relation. In the case when the preference relation is assumed to be monotone, all life situations that dominate a hypothetical life situation that is preferred over the current life situation, also belong to the upper contour set. Likewise, all life situations that are dominated by a hypothetical life situation over which the current life situation is preferred, belong to the lower contour set. In the case when the preference relation is assumed to be convex as well, the upper contour set also contains all life situations that can be obtained as a convex combination of two life situations which are already in the upper contour set. These observations allow us to chart sets, which are subsets of the true upper and lower contour set of the respondent. Because only a small number of steps of the bisectional algorithm is used in each set of dichotomous choices, these charted sets underestimate the true size of the upper and lower contour set of the respondents. Yet, we will illustrate below that our empirical estimates are rather "tight" and provide narrow bounds on the non-comparable sets. Based on these bounds, the extended upper and lower contour set can

be derived easily using expression (1) and (2). The obtained estimates of the extended upper and lower contour set are conservative estimates in the sense that they may overestimate their size, which makes it harder for the *Nested Contours* criterion to be fulfilled.

The right-hand panel of Figure 4 illustrates the second step of the ABDC method. In the first step, the bisectional algorithm has been used in four sets of dichotomous choices, which are indicated in the right-hand panel of Figure 4 by means of dashed lines. The bisectional algorithm in each set leads to a pair of life situations that bound the non-comparable set, indicated by the circles in Figure 4. When monotonicity of the preference relation is assumed, a conservative estimate of the upper and lower contour set is illustrated in dark grey in the figure. When the preference relation is assumed to be convex as well, the conservative estimate of the upper contour contains the light grey area as well. The method illustrated in Figure 4 is repeated for each respondent in our data set and provides us with conservative estimates of the individual extended upper and lower contour set in an entirely non-parametric way.

5 Results

We present two sets of results in this section. First, we study the extent to which responses in the first step of the ABDC method are consistent with some standard (parametric) properties of a preference relation, such as monotonicity, convexity, linearity or having a constant elasticity of substitution. Second, we implement the E_λ-*Restricted Nested Contours* criterion for all pairs of respondents in our data set and quantify the robustness of their pairwise ranking by determining the maximal λ for which the pair of respondents can be ranked.

5.1 Properties of the preference relations

Since each dichotomous choice in the ABDC method involves the current life situation and only one hypothetical life situation, it is impossible to test the consistency of the responses with transitivity directly. We can, however, test consistency of the preference relation with transitivity in combination with other properties, such as monotonicity and convexity. In fact, our data set allows doing that in two different ways.

For the first test, we have included an additional fifteenth dichotomous choice among the list of choices that each respondent makes. This additional choice is included between the first and the second set of choices. The hypothetical life situation in this additional choice is chosen either to dominate or to be dominated by the hypothetical life situation from the first dichotomous choice based on the choice of the respondent (see Appendix 5 for more details). By monotonicity and transitivity, the respondent should make the same choice in this test as in the first dichotomous choice. The first row of Table 2 shows the results of this consistency check. We see that 17.5% of the respondents do not pass this test in the baseline.

How much is a fail rate of 17.5%? It is hard to interpret this number without any reference about the severity or power of the test. To provide such a reference, we consider the following thought experiment. For the experiment we construct an artificial data set in which all dichotomous choices are made in a purely random manner. For each choice, the options "own life situation", "hypothetical life situation", and "I don't know" are selected with probability one third. This artificial data set mimics a world in which respondents have no consistent preferences or are totally disinterested in responding truthfully to the survey.[17] The second column of Table 2 gives the results of the consistency test in this thought experiment.[18] About 52.9% of the artificial respondents fail the test, which is considerably more than in our data set.[19]

For the second test, the ABDC method provides an alternative way of testing whether the responses are consistent with transitivity and monotonicity of the preference relation. At least one of the hypothetical life situations that is considered in the set of dichotomous choices with an intermediate reference level for health or consumption is vector dominated by a hypothetical life situation from the set of choices where the reference level is fixed at the maximal level. If a respondent with a transitive and monotone preference relation prefers her current

[17] On the contrary, answers in an open feedback field at the end of the survey indicate that at least some respondents were interested in the survey and enjoyed their participation: "Survey was very interesting and I enjoyed it", "I liked it, it made me think about what's really important in life" or "Weird survey. Very interesting. Thanks for allowing to participate.". Yet, some participants found that "some questions were a little confusing", while a small minority of respondents seemed to dislike the survey "I think it was a very stupid survey" or its purpose "trying to quantify intangibles is iffy at best".

[18] Selten (1991) and Beatty and Crawford (2011) suggest to look at the difference between the fail rate in the random data set and the fail rate of the actual data set as a measure of predictive success.

[19] As discussed in Appendix 5, five out of the nine possible combinations of responses to the first dichotomous choice and the consistency test fail the test.

Table 2: Fail rates of consistency tests on properties of preference relations

	baseline ABDC	random data	less choices	less sets
Test 1 (additional choice)	17.5	52.9	17.5	17.5
Test 2 (transitivity and monotonicity)	12.0	41.6	8.7	0
Test 3 (transitivity and convexity)	9.7	42.2	6.9	1.1
Test 4 (CES preferences)	32.5	83.5	22.9	1.0
Test 5 (linear preferences)	95.4	98.0	92.7	88.1
Test 6 (Cobb Douglas preferences)	71.9	96.0	63.3	63.4
Test 7 (Leontief preferences)	48.8	95.9	36.7	28.8
Test 8 (Kinked linear preferences)	70.0	86.6	61.8	0
N	2,260	2,260	2,260	2,260

life situation over the latter hypothetical life situation, then she should make the same choice when considering the dominated hypothetical life situation. Respondents fail our second test if they prefer the dominated hypothetical life situation over their current life situation. The left-hand panel of Figure 6 shows such a situation when ℓ_i is preferred over the dominating life situation B, while the dominated life situation C is preferred over ℓ_i. The second row of Table 2 indicates that these situations occur for 12.0% of the respondents, which is a considerably lower fail rate than in the thought experiment with purely random responses, as can be seen in the second column of Table 2.

Similarly, the third test considers whether the responses are consistent with transitivity and convexity of the preference relation. These properties together imply that life situations which belong to the convex hull of life situations that are preferred over the current life situation, should also be preferred. The test checks precisely that. The right-hand panel illustrates a case when the test fails if life situation A is preferred over ℓ_i and ℓ_i is preferred over life situation F. The third row of Table 2 presents the fail rate of this test. About 9.7% of the respondents fail the test, which is again considerably lower than in the thought experiment with purely random responses.

Column three and four provide sensitivity checks for the reported fail rates. Both sensitivity checks look at the fail rates when less information from the ABDC method is used. For the third column, the answers to the last dichotomous choice in each of the four sets is discarded. Since information about less choices

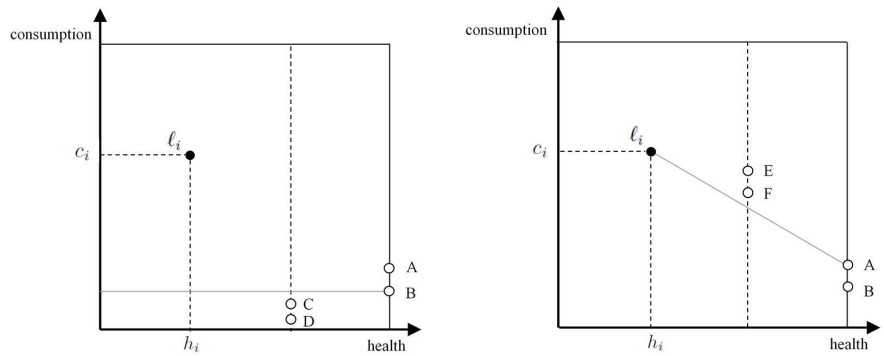

Figure 6: The left-hand panel shows a situation in which Test 2 (transitivity and monotonicity) fails. The right-hand panel shows a situation in which Test 3 (transitivity and convexity) fails.

is used, the non-comparable set of each respondent is larger and the second and third tests have less power, which results in lower fail rates. The size of this drop in fail rates informs us about the usefulness of including the last dichotomous choice in each of the four sets. As every choice tightens the bounds around the indifference curves, we get an idea of the role plaid by the tightness of the bounds on the results by comparing the third and first column. The results in the fourth column discard the information gathered in the sets of dichotomous choices with an intermediate reference health or consumption level. Without these sets of dichotomous choices, the second test has no power at all, while the third test has very little power.

Moreover, the ABDC method also permits to test parametric specifications that are commonly imposed on preference relations or on a utility function that represents it. Consider, for instance, the constant elasticity of substitution (CES) utility function

$$U(c_i, h_i) = (w \times c_i^{\gamma} + (1 - w) \times h_i^{\gamma})^{1/\gamma}, \tag{4}$$

where w is the relative weight assigned to consumption and γ a parameter capturing the curvature of the implied indifference curves. The parameter γ equals $(\sigma - 1)/\sigma$, where σ is the elasticity of substitution between the dimensions. All preference relations that can be represented by a CES utility function, are

homothetic. The CES specification is generally perceived as a flexible one.[20] If the parameter γ equals 1, the utility function is linear with perfect substitution between consumption and health. If $\gamma \to 0$, we obtain the Cobb-Douglas utility function and if $\gamma \to -\infty$, we get the Leontief or perfect complements utility function.

The bottom part of Table 2 tests whether the individual responses provided by the ABDC method can be rationalized by a CES utility function with individual-specific parameters. The test examines for each respondent whether there exists an indifference curve of a CES utility function that fits entirely in the non-comparable set $NC(\ell_i, P_i)$. In other words, she passes the test when the indifference curve of the considered utility function through the current life situation is such that all hypothetical life situations that are preferred to the current life situation lie above the indifference curve, and all hypothetical life situations over which the current life situation is preferred lie below the curve. Appendix 6 presents the algorithm that is used to implement the test.

The fourth row of Table 2 considers the broad class of general CES utility functions as described by equation (4). We find that 32.5% of all respondents fail the test, while 83.5% of the artificial respondents in the random data do the same. The difference is remarkably large, which indicates that a general CES utility function does a reasonable job to rationalize the observed data, whereas it would perform much worse for the artificial random data. The sensitivity check in column 4 of Table 2 shows that the test becomes virtually powerless without the sets of dichotomous choices with an intermediate reference level.

As the next three rows of the Table 2 show, fixing the curvature parameter γ at a specific value increases the fail rate dramatically, especially for the linear case. In short, the choices made by almost all (95.4%) respondents in the ABDC method cannot be rationalized with a linear utility function. This finding raises fundamental questions about the widespread use of linear utility functions to model choices in similar settings, for instance in discrete choice experiments to determine the WTP for health improvements (see Bateman et al. (2002) for a survey). The results for the Cobb Douglas and, in particular, the Leontief preferences show lower fail rates than the linear case, indicating that the responses of many respondents are better rationalized by an indifference curve

[20] Virtually all composite indices of well-being in the literature fit in this parametric specification, see Decancq and Lugo (2013) for an overview. Blackorby and Donaldson (1982) provide an axiomatic characterisation, and Maasoumi (1986) an information theoretic justification for a CES well-being index.

with a considerable curvature.

Triggered by the large fail rate of the linear specification and inspired by the wide body of experimental evidence on the endowment effect in the wake of Kahneman et al. (1990), we finally check whether a "kinked" linear utility function with a kink at the current life situation would rationalize the data better than a straight linear utility function. A "kinked" linear utility function captures the finding that the "willingness-to-accept" a worsening of health outcomes is typically found to exceed the "willingness-to-pay" for improvements in health. We see on the ultimate row of Table 2 that the fail rate of the "kinked" linear utility function is indeed considerably lower than the one of the straight linear utility function (70.0% versus 95.4%), yet a comparison with the results for the artificial data in the second column shows that the test is not very powerful either.[21]

5.2 Interpersonal well-being comparisons

We now turn to the central question of the paper and measure the power of the proposed criteria to make interpersonal well-being comparisons. To determine whether two respondents can be ranked with the proposed E_λ-*Restricted Nested Contours* criterion, we first need to make a choice for a specific sequence of nested evaluation sets $\mathcal{E}$. We consider the sequence that consists of the following evaluation sets:

$$E_\lambda = \{\ell = (c, h) \in \mathcal{L} | h \geq 100 - \lambda\} \text{ for each } \lambda \in [0, 100]\,. \tag{5}$$

These evaluation sets are indeed nested with $E_\lambda \subset E_{\lambda'}$ whenever $\lambda < \lambda'$. The maximal and minimal values of λ provide natural benchmark evaluation sets. When $\lambda = 100$, the evaluation set equals the entire space of life situations $\mathcal{L}$, and E_λ-*Restricted Nested Contours* is equivalent to *Nested Contours*. When $\lambda = 0$, the evaluation set contains all life situations with a maximal health level. The E_0-*Restricted Nested Contours* criterion requires that the intersection between the extended upper and lower contour set is empty for those life situations in

[21] In our test, all indifference curves of the "kinked" linear utility function are required to have their kink at the current life situation, whereas the indifference curves of the Leontief preferences may have their kink elsewhere. That explains why the fail rate of the Leontief preferences is lower in the first column of Table 2 than the fail rate of the "kinked" linear utility function.

which the health level is maximal.[22]

Equipped with an estimate of the extended upper and lower contour set for each respondent and a sequence of nested evaluation sets, we test whether the well-being of two respondents can be compared with E_λ-*Restricted Nested Contours*. We proceed as follows: for each λ value we check for all pairs of respondents whether they can be ranked according to the E_λ-*Restricted Nested Contours* criterion. Figure 3 reports for each λ value the percentage of pairs of respondents that can be ranked. We consider two cases in the figure. First, the case when the preference relation is assumed to be monotonic (the full line in the figure) and, second, the case when the preference relation is also assumed to be convex (the dashed line). In these pairwise comparisons all respondents whose preferences fail to satisfy the considered properties have been dropped. Nevertheless, many pairs of respondents remain to be compared (respectively 3,950,156 and 3,310,580 pairs).

Let us start with the results when only monotonicity of the preference relation is assumed. Figure 3 shows that 51.1% of all pairs can be compared for the evaluation set E_0, i.e., the evaluation set that consists of all life situations for which health is maximal. Increasing the value of the parameter λ, increases the evaluation set, so that it becomes harder to fulfil the criterion. We see that the percentage of pairs that can be ranked reduces gradually, until it plateaus around 20% for λ values larger than 50. In other words, we find that one fifth of all pairs can be ranked, even when using the most restrictive version of the criterion. The additional assumption of convexity of the preference relation provides us with larger estimates of the upper contour sets of the respondents and, hence, a smaller estimate of its complement, the extended lower contour set LC^+, so that more pairs can be ranked. In practice, the number of extra pairs that can be ranked is rather small, between 3 and 4%, as can be seen from comparing both curves in Figure 3. Table 3 provides more details and presents a sensitivity analysis for the alternative cases when less dichotomous choices are considered in each set (columns 3 and 4) or when less sets are considered

[22] When preferences are complete, well-being comparisons based on E_0-*Restricted Nested Contours* are based on a comparison of the full-health equivalent incomes of the individuals (see Fleurbaey and Schokkaert (2009) and Decancq et al. (2015a,b)). In this case, the criterion leads to a complete ranking. When preferences are incomplete, on the other hand, the proposed criterion is a special case of the criterion implied by Proposition 3 of Fleurbaey and Schokkaert (2013). Due to the incompleteness of the preferences, the criterion leads to an incomplete ranking. Fleurbaey and Schokkaert propose therefore further restrictions based on so-called "safety principles" to make the well-being comparisons more complete.

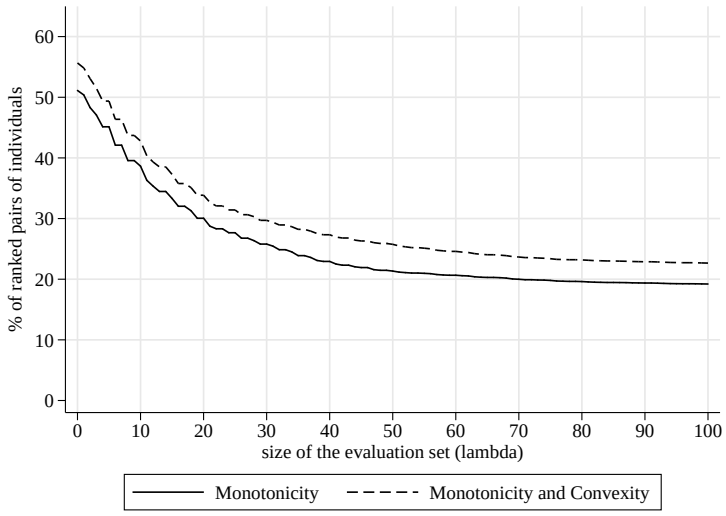

Figure 7: Percentage of ranked pairs of individuals with the E_λ-*Restricted Nested Contours* criterion, by size of the evaluation set λ

(columns 5 and 6). The general trend of the results is very similar, with a larger gain from convexity when less sets are considered. A comparison between the third and first column reveals that the tightening of the bounds by asking an additional choice in each set, allows us to rank between 3% and 5% additional pairs.[23]

Finally, we look at some socio-demographic characteristics of the respondents who are considered better-off according to *Nested Contours* in our data set. To do that, we count for each respondent the number of other respondents who are worse-off. We say that the more respondents are considered worse off, the better off the respondent at hand is. While many respondents (60.2% to be precise) are not better off than any other respondent, some respondents are better off than many other respondents.[24] Figure 8 shows the average number of respondents who are considered as worse off, broken down by ethnic group

[23] An alternative incomplete criterion for well-being comparisons based on vector dominance (advocated by Sen (1985) and discussed by Brun and Tungodden (2004)), which states that $(\ell_i, P_i) \succ (\ell_j, P_j)$ when $\ell_i > \ell_j$, ranks 54.8% of all pairs of respondents in our data set. On average, respondents dominate 638 respondents from a data set of 2,260 respondents. There is a respondent who dominates not less than 2248 other respondents.

[24] One respondent is better off than no less than 1,438 other respondents, about 63.6% of the entire data set. She is a white woman with a college degree, a health index of 98, and a consumption level of $2,743.

Table 3: Percentage of ranked pairs of individuals with the E_λ-*Restricted Nested Contours* criterion, by size of the evaluation set λ

λ	Mon. (baseline)	Mon. and Conv. (baseline)	Mon. (less choices)	Mon. and Conv. (less choices)	Mon. (less sets)	Mon. and Conv. (less sets)
0	51.1	55.7	46.3	50.8	30.8	59.8
10	38.6	42.8	35.3	39.1	25.3	40.0
20	30.1	33.9	27.2	30.6	20.4	30.8
30	25.8	29.7	21.7	25.2	18.2	26.8
40	23.0	27.3	18.6	22.4	16.9	24.7
50	21.3	25.8	17.1	20.8	16.3	23.6
60	20.7	24.6	16.3	19.6	15.5	22.6
70	20.0	23.7	15.5	18.6	14.8	21.7
80	19.6	23.1	15.1	18.1	14.3	21.2
90	19.4	22.9	14.8	17.7	14.0	20.8
100	19.2	22.7	14.6	17.5	13.9	20.7
N	3950156	3310580	4258032	3753906	5105340	4997460

(on the left) and highest educational attainment (on the right). In the left-hand panel, it can be seen that Asian Americans and, surprisingly, Native Americans are better off, whereas Black Americans and Latinos, are found to be worse off on average. Moreover, in the right-hand panel, there appears to be a clear educational gradient in average number of respondents who are considered as worse off.[25]

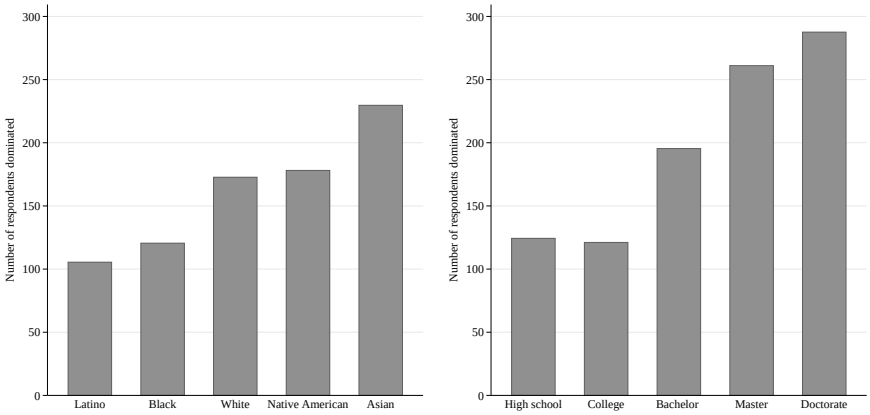

Figure 8: Average number of respondents considered to be worse off by the *Nested Contours* criterion in our sample, by ethnicity (left-hand panel) and highest degree attained (right-hand panel)

6 Conclusion

We have investigated how powerful *Nested Contours* is to make real-world interpersonal well-being comparisons. The criterion formalizes an intuition which is natural in an approach that builds on the idea that well-being comparisons should respect the ordinal preferences of the concerned individuals. It states that one individual is better off than another one if the intersection between the extended upper contour set of the better off individual and the extended lower contour set of the worse off individual is empty. The criterion can be generalized in a flexible way by E_λ-*Restricted Nested Contours*, which requires the

[25] Note, however, that these results are mainly driven by the higher consumption and health level of these groups. Table 8 of Appendix 7 shows that ethnicity and educational attainments have little explanatory power in an OLS regression with the number of worse-off respondents as explained variable in presence of consumption and health outcomes of the respondents.

emptiness of the intersection between the contour sets on an evaluation set E_λ only. The size of the largest evaluation set for which E_λ-*Restricted Nested Contours* holds, provides a natural measure of the robustness of the interpersonal well-being comparison between two individuals. To implement the criterion, the extended upper and lower contour set of both concerned individuals needs to be known. We have proposed the ABDC method to chart these contours sets in an entirely non-parametric way in the consumption-health space. The method is based on a small number of dichotomous choices and mild non-parametric assumptions on the individual preferences.

In an online sample with 2,260 American respondents, we have found that the ABDC method provides relatively tight bounds on the individual extended upper and lower contour sets, despite our conservative modelling approach that allows respondents to report (potentially very) incomplete information about their preferences. The obtained contour sets have been shown to be consistent with our non-parametric assumptions of monotonicity and convexity for a large majority of respondents, while common parametric assumptions on preference relations are rejected for an overwhelming number of respondents. About one fifth of all pairs of respondents can be ranked with *Nested Contours* alone, while a more complete version of E_λ-*Restricted Nested Contours* is able to rank more than half of all pairs of respondents. Asian American and highly educated Americans are found to be better off than a larger number of other respondents, compared to lower educated Black Americans or Latinos. It is an open question how the results extend to other countries or to different settings with other dimensions of well-being.

In our theoretical and empirical analysis, we have not imposed the individual preference relation to be complete and we have chosen to remain agnostic about the exact source of this incompleteness. Either the individuals are unable to rank all life situations, or the observer may only manage to observe incomplete information on the preferences from the elicitation process. One can argue that both sources of incompleteness deserve a different treatment from a normative perspective. It would therefore be interesting to try to disentangle both sources of incompleteness in follow-up studies.

Although already a large number of pairs of individuals can be ranked with the incomplete *Nested Contours* criterion alone, designing and evaluating policies requires presumably a complete ranking of all individuals. Additional, stronger and arguably less appealing criteria are needed for such a complete ranking. The

literature contains many examples of how *Nested Contours* can be strengthened, leading to money metrics utilities, ray utilities or equivalent income measures as well-being measures (see Fleurbaey and Maniquet (2011, 2017)).

Clearly, the ABDC method to elicit preferences over non-market goods is still in its infancy. We leave questions about its implementation in a setting with a large number of dimensions, respondents who answer strategically, or its consistency with real-world choices as interesting avenues for further research.

Acknowledgements

We thank Bart Capéau, Laurens Cherchye, Marion Collewet, Angus Deaton, Bruno De Borger, Thomas Demuynck, Bram De Rock, Marc Fleurbaey, François Maniquet, Erwin Ooghe, Erik Schokkaert, Dirk Van de gaer, Frederic Vermeulen, Verity Watson, and participants at seminars in Princeton, Louvain-la-Neuve, LSE, Aberdeen, Oxford, Leuven, Antwerp and Umeå for helpful discussions and comments. Financial support of the Research Foundation - Flanders is gratefully acknowledged. This paper was partly developed when Koen Decancq was a Visiting Research Scholar at the Center for Health and Wellbeing of Princeton University, and he thanks the Center for providing support.

Appendix 1. Summary statistics of the data set

Table 4: Summary Statistics

Variable	N	Mean	Std. Dev.
Consumption	2,260	883.4	633.3
Health index	2,260	64.1	27.8
Male	2,260	48.8	50.0
White	2,260	82.5	38.0
Latino	2,260	4.1	19.9
Black	2,260	7.7	26.6
Native American	2,260	1.1	10.4
Asian	2,260	3.2	17.6
Age between 25 and 34	2,260	20.2	40.2
Age between 35 and 44	2,260	17.3	37.8
Age between 45 and 54	2,260	19.1	39.3
Age between 55 and 64	2,260	24.9	43.2
Age between 65 and 74	2,260	16.0	36.5
Age between 75 and older	2,260	2.7	16.3
High school	2,260	25.9	43.6
College	2,260	23.6	42.5
Bachelor's degree	2,260	38.9	48.8
Master's degree	2,260	9.7	29.5
Doctorate's degree	2,260	1.9	13.6

Appendix 2. Consumption items

Household expenditures	N	Mean	Std. dev.	Min.	Max.
Trips and holidays	2,260	69.6	266.1	0	5,000
Children	2,260	45.2	134.8	0	2,000
Transport	2,260	86.8	123.8	0	2,000
Food indoors	2,260	275.2	249.3	0	3,000
Housekeeping	2,260	28.8	79.4	0	2,559
Other	2,260	51.7	162.3	0	3,000
Personal expenditures	N	Mean	Std. dev.	Min.	Max.
Food outdoors	2,260	68.3	112.6	0	2,520
Cigarettes	2,260	23.1	59.1	0	985
Clothing	2,260	42.0	79.3	0	1,000
Personal care	2,260	40.9	77.3	0	1,000
Leisure	2,260	29.4	58.8	0	1,000
Schooling	2,260	6.7	44.6	0	1,000
Donations	2,260	25.2	74.4	0	1,200
Other personal expenses	2,260	25.4	74.8	0	1,000

Table 5: Consumption items

Appendix 3. Health items

Sub dimension	Items
Physical health	1. Does your health limit you in moderate activities? 2. Does your health limit you when climbing several flights of stairs? 3. Did you accomplish less as a result of physical health during the past four weeks? 4. Have you been limited as a result of physical health during the past four weeks? 5. How much did pain interfere with your normal work during the past four weeks?
Emotional health	6. Did you accomplish less as a result of emotional health during the past four weeks? 7. Did you do your work less carefully than usual as a result of emotional health during the past four weeks? 8. Have you felt calm and peaceful during the past four weeks? 9. Did you have a lot of energy during the past four weeks? 10. Have you felt down-hearted and blue during the past four weeks?

Table 6: Items of the SF-12 used to construct the health sub dimensions

Appendix 4. The dichotomous choices

Figure 9 provides a screenshot of a dichotomous choice that a respondent with a consumption of $470 and a health level of 82 is confronted with.

Imagine that both situations are possible for you, which one would you unambiguously choose?
If you cannot make an unambiguous choice between both situations, you can tick the response "I don't know".

Recall that both situations are assumed to be **identical in all other aspects.**

	YOUR LIFE
Consumption	$470
Health	82%

I DON'T KNOW

	ANOTHER LIFE
Consumption	$235
Health	91%

Figure 9: Screenshot of a dichotomous choice

To help respondents understanding the meaning of the health level used in the hypothetical life situations, we construct 16 vignettes. Each vignette describes a possible set of responses to the 10 items listed in Table 6 that correspond to the hypothetical health level. As example, Figure 10 provides a screenshot of the vignette corresponding to a hypothetical health level between 88 and 94.

For your information, the following situation corresponds to the health level described in "ANOTHER LIFE" in the table above:

During the last 4 weeks, as a result of **physical problems**,

- you didn't accomplish less than you would like;
- you weren't limited in the kind of work and other activities;
- physical pain interfered moderately in your normal work;
- you were not limited at all in moderate activities such as moving a table, pushing a vacuum cleaner or bowling;
- you were not limited at all in activities such as climbing several flights of stairs.

During the last 4 weeks, as a result of **emotional problems**,

- you didn't accomplish less than you would like;
- you did your work or activities as carefully as usual;
- you felt peaceful and calm most of the time;
- you felt downhearted and blue a little of the time;
- most of the time you had a lot of energy.

Figure 10: Screenshot of the corresponding vignette

Appendix 5. Test 1 (additional choice)

The fifth dichotomous choice that the respondents make is a consistency test. This test is constructed based on the answers of the first dichotomous choice. Three cases can be distinguished based on their initial choice.

First, if the respondent preferred her own life situation in the first dichotomous choice, the hypothetical choice of the consistency test is chosen so that it is vector dominated by the hypothetical choice of the first choice. A respondent with transitive and monotone preferences then also prefers her own life situation in the consistency test.

Second, if the respondent preferred the hypothetical life situation in the first dichotomous choice, the hypothetical choice of the consistency test is chosen so that it vector dominates the hypothetical choice of the first choice. A respondent with transitive and monotone preferences then also prefers the hypothetical life situation in the consistency test.

Third, if the respondent didn't know whether she preferred her own life situation or the hypothetical life situation in the first dichotomous choice, the hypothetical choice of the consistency test is chosen so that it vector dominates the hypothetical choice of the first choice. A respondent with transitive and monotone preferences does not prefer her own life situation in the consistency test. Indeed, both the choice for the hypothetical life situation and an answer "I don't know" are compatible with transitivity and monotonicity.

Table 7 presents the results of the consistency test in our data set. In the rows of the table, we tabulate the answers to the first dichotomous choice. We see that 618 respondents have chosen their own life situation, for instance. In the columns, we tabulate the answers to the consistency test. The answers in a bold face pass the test, the others do not. We see that 17.5% of the respondents does not pass the test.

	own life sit.	hypothetical	"I don't know"	
own life sit.	**513**	113	55	618
hypothetical	86	**927**	69	1,082
"I don't know"	72	**136**	**289**	497
	671	1,176	413	2,260

Table 7: Results of the consistency test in the baseline

Appendix 6. Test of CES utility specification

To test whether a CES utility function (expression (4)) is able to rationalize the dichotomous choices made by a respondent in the ABDC method, we check whether there exists an indifference curve of this utility function such that all preferred life situations lie above the indifference curve and all life situations to which the current life situation is preferred set lie below. Indifference curves of a CES utility function satisfy the following equation

$$c^\gamma = \alpha - \beta \times h^\gamma \tag{6}$$

for some parameters α, β, and γ. We use the following algorithm to perform the test.

Algorithm. *The algorithm proceeds on a grid of γ and c_2 values. For each γ, the parameters α and β are determined such that life situations (c_1, h_1) and (c_2, h_2) belong to an indifference curve with specification (6). The life situation (c_1, h_1) is the current life situation of the respondent. The life situation (c_2, h_2) is chosen on a second grid such that h_2 equals full health and c_2 is a consumption level between the lower and upper bound indicated by the respondent in the ABDC method in the first set of dichotomous choices.*[26]

The parameters α and β are obtained by solving a system of two equations for a given γ value and life situations (c_1, h_1) and (c_2, h_2)

$$\begin{cases} c_1^\gamma = \alpha - \beta \times h_1^\gamma \\ c_2^\gamma = \alpha - \beta \times h_2^\gamma. \end{cases} \tag{7}$$

This yields:

$$\alpha = \frac{h_1^\gamma c_2^\gamma - c_1^\gamma h_2^\gamma}{h_1^\gamma - h_2^\gamma} \text{ and } \beta = -\frac{c_1^\gamma - c_2^\gamma}{h_1^\gamma - h_2^\gamma}. \tag{8}$$

Once α and β are computed, it is tested whether all preferred life situations lie above the indifference curve and all life situations to which the current life situation is preferred set lie below the curve. If that is the case, the algorithm proceeds to the next value on the grid until the last value is reached. If that is not the case, the respondent passes the test and the algorithm stops.

[26] The reported results are obtained with an equally spaced grid of 160 γ values between -15 and 1 and 21 equally spaced values for c_2 between the indicated lower and upper bound.

Appendix 7. Determinants of the number of respondents that are worse-off

	Model 1		Model 2	
Constant	-0.048***	(0.013)	-0.060**	(0.024)
Consumption/1000	0.066***	(0.008)	0.062***	(0.008)
Health index/10	0.029***	(0.002)	0.028***	(0.002)
College			-0.023	(0.014)
Bachelor's degree			0.006	(0.013)
Master's degree			0.035*	(0.019)
Doctorate's degree			0.066*	(0.037)
Latino			-0.021	(0.025)
Black			-0.025	(0.018)
Native American			0.040	(0.046)
Asian			0.019	(0.028)
Age			0.000	(0.000)
Female			-0.011	(0.010)
N	2,260		2,185	
R^2	0.149		0.159	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Determinants of the number of respondents that are worse-off

References

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2006): "Elicitation using multiple price list formats," *Experimental Economics*, 9, 383–405.

Arrow, K. (1950): "A difficulty in the concept of social welfare," *Journal of Political Economy*, 58, 328–346.

——— (1973): "Some ordinalist-utilitarian notes on Rawls' theory of justice," *Journal of Philosophy*, 70, 245–263.

Bateman, I. J., R. Carson, B. Day, M. Hanemann, N. Hanley, T. Hett, and M. Jones-Lee (2002): *Economic Valuation with Stated Preference Techniques: A Manual.*, Cheltenham and Northampton: Edward Elgar.

Beatty, T. K. and I. A. Crawford (2011): "How demanding is the revealed preference approach to demand?" *The American Economic Review*, 101, 2782–2795.

Bernheim, B. D. and A. Rangel (2009): "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics," *The Quarterly Journal of Economics*, 124, 51–104.

Blackorby, C. and D. Donaldson (1982): "Ratio-scale and translation-scale full interpersonal comparability without domain restrictions: Admissible social-evaluation functions," *International Economic Review*, 23, 249–68.

Bosmans, K., K. Decancq, and E. Ooghe (2018): "Who is affraid of aggregating money metrics?" *Theoretical Economics*, 13, 467–484.

Brun, B. C. and B. Tungodden (2004): "Non-welfaristic theories of justice: Is "the intersection approach" a solution to the indexing impasse?" *Social Choice and Welfare*, 22, 49–60, 10.1007/s00355-003-0276-0.

Butler, D. and G. Loomes (1988): "Decision difficulty and imprecise preferences," *Acta Psychologica*, 68, 183–196.

——— (2007): "Imprecision as an account of the preference reversal phenomenon," *The American Economic Review*, 97, 277–297.

Cherchye, L., B. De Rock, and F. Vermeulen (2012): "Married with children: A collective labor supply model with detailed time use and intra-household expenditure information," *The American Economic Review*, 102, 3377–3405.

Deaton, A. (1979): "The distance function in consumer behaviour with applications to index numbers and optimal taxation," *The Review of Economic Studies*, 46, 391–405.

Decancq, K., M. Fleurbaey, and F. Maniquet (2014): "Multidimensional poverty measurement with individual preferences," Tech. rep., Princeton University - William S. Dietrich II Economic Theory Center Working Paper 058-2014.

Decancq, K., M. Fleurbaey, and E. Schokkaert (2015a): "Happiness, equivalent incomes, and respect for individual preferences," *Economica*, 82, 1082–1106.

——— (2015b): "Inequality, Income, and Well-Being," in *Handbook of Income Distribution*, ed. by A. B. Atkinson and F. Bourguignon, Elsevier, vol. 2 of *Handbook of Income Distribution*, 67 – 140.

Decancq, K. and M. A. Lugo (2013): "Weights in Multidimensional Indices of Well-Being: An Overview," *Econometric Reviews*, 32, 7–34.

Dolan, P. and D. Kahneman (2008): "Interpretations of utility and their implications for the valuation of health," *The Economic Journal*, 118, 215–234.

Fleurbaey, M. and D. Blanchet (2013): *Beyond GDP: Measuring Welfare and Assessing Sustainability*, Oxford University Press.

Fleurbaey, M. and F. Maniquet (2011): *A Theory of Fairness and Social Welfare*, Cambridge University Press.

——— (2017): "Fairness and well-being measurement," *Mathematical Social Sciences*, 90, 119–126.

Fleurbaey, M. and E. Schokkaert (2009): "Unfair inequalities in health and health care," *Journal of Health Economics*, 28, 73–90.

——— (2013): "Behavioral Welfare Economics and Redistribution," *American Economic Journal: Microeconomics*, 5, 180–205.

Fleurbaey, M. and K. Tadenuma (2014): "Universal social orderings: An integrated theory of policy evaluation, inter-society comparisons, and interpersonal comparisons," *The Review of Economic Studies*, 81, 1071–1101.

Hammond, P. J. (1991): "Interpersonal comparisons of utility: Why and how they are and should be made," in *Interpersonal Comparisons of Well-Being*, ed. by J. E. R. Jon Elster, Cambridge University Press London, 200–254.

Hansson, B. (1973): "The independence condition in the theory of social choice," *Theory and Decision*, 4, 25–49.

Hardisty, D. J., K. F. Thompson, D. H. Krantz, and E. U. Weber (2013): "How to measure time preferences: An experimental comparison of three methods," *Judgment and Decision Making*, 8, 236.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1990): "Experimental tests of the endowment effect and the Coase theorem," *Journal of Political Economy*, 98, 1325–1348.

King, M. A. (1983): "Welfare analysis of tax reforms using household data," *Journal of Public Economics*, 21, 183 – 214.

Le Lec, F. and M. Krawczyk (2018): "Eliciting distributional preferences," Tech. rep., Mimeo.

Loomes, G., C. Starmer, and R. Sugden (1991): "Observing violations of transitivity by experimental methods," *Econometrica*, 425–439.

Maasoumi, E. (1986): "The measurement and decomposition of multidimensional inequality," *Econometrica*, 54, 991–997.

MacCrimmon, K. R. and M. Toda (1969): "The experimental determination of indifference curves," *The Review of Economic Studies*, 36, 433–451.

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

McKenzie, L. (1957): "Demand theory without a utility index," *The Review of Economic Studies*, 24, 185–189.

Moscati, I. (2007): "Early experiments in consumer demand theory: 1930-1970," *History of Political Economy*, 39, 359–402.

Pazner, E. A. (1979): "Equity, nonfeasible alternatives and social choice: A reconsideration of the concept of social welfare," in *Aggregation and Revelation of Preferences*, ed. by J. Laffont, Amsterdam: North-Holland.

Prades, J. L. P., I. M. Perez, and G. Loomes (2014): "Comparing iterative choice-based methods in the elicitation of utilities for health states," Tech. rep.

Robbins, L. (1932 [1984]): *An essay on the nature and significance of economic science*, New York: New York University Press.

——— (1938): "Interpersonal Comparisons of Utility: A Comment," *The Economic Journal*, 48, 635–641.

Salant, Y. and A. Rubinstein (2008): "(A, f): choice with frames," *The Review of Economic Studies*, 75, 1287–1296.

Samuelson, P. A. (1948): "Consumption theory in terms of revealed preference," *Economica*, 15, 243–253.

——— (1974): "Complementarity: an essay on the 40th anniversary of the Hicks-Allen revolution in demand theory," *Journal of Economic Literature*, 12, pp. 1255–1289.

——— (1977): "Reaffirming the existence of "reasonable" Bergson-Samuelson social welfare functions," *Economica*, 81–88.

Selten, R. (1991): "Properties of a measure of predictive success," *Mathematical Social Sciences*, 21, 153–167.

Sen, A. K. (1985): *Commodities and Capabilities*, Amsterdam and Oxford: North-Holland.

Stiglitz, J. E., A. Sen, and J.-P. Fitoussi (2009): "Report by the Commission on the measurement of economic performance and social progress," Mimeo.

Thurstone, L. L. (1931): "The indifference function," *The Journal of Social Psychology*, 2, 139–167.

Toubia, O., E. Johnson, T. Evgeniou, and P. Delquié (2013): "Dynamic experiments for estimating preferences: An adaptive method of eliciting time and risk parameters," *Management Science*, 59, 613–640.

Tversky, A. (1969): "Intransitivity of preferences." *Psychological Review*, 76, 31.

Ubel, P. A., G. Loewenstein, and C. Jepson (2005): "Disability and sunshine: can hedonic predictions be improved by drawing attention to focusing illusions or emotional adaptation?" *Journal of Experimental Psychology: Applied*, 11, 111.

Varian, H. R. (1982): "The nonparametric approach to demand analysis," *Econometrica*, 945–973.

Ware, J. E. J., M. Kosinski, and S. D. Keller (1996): "A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity," *Medical Care*, 34, 220–233.

Watson, V. and M. Ryan (2007): "Exploring preference anomalies in double bounded contingent valuation," *Journal of Health Economics*, 26, 463–482.

Weisbrod, B. A. (1977): "Comparing utility functions in efficiency terms or, what kind of utility functions do we want?" *The American Economic Review*, 67, 991–995.