University of Antwerp
AIPRIL | Centre of Excellence

Incomplete Preferences, Well-Being Measurement, and the Identification of the Worst-Off

Santiago Burone & Koen Decancq

Working Paper

No. 26/01

February 2026

CSB Centrum voor Sociaal Beleid Herman Deleeck Universiteit Antwerpen
CSG Centre for Urban History University of Antwerp
IOB Institute of Development Policy University of Antwerp

Incomplete Preferences, Well-Being Measurement, and the Identification of the Worst-Off

Santiago Burone[1] Koen Decancq[1,2]

[1] Herman Deleeck Centre for Social Policy, University of Antwerp

[2] Department of Economics (KULeuven)

Working Paper No. 26/01

February 2026

Abstract

When individual preferences are incomplete, the information available for well-being measurement is interval-valued rather than point-identified. We show that in this case, well-being measurement involves an unavoidable normative choice: any procedure that delivers policy-relevant complete rankings must resolve incomparability in a substantive way. We axiomatically characterize well-being measures under incomplete preferences and show that any measure satisfying four natural axioms must aggregate the bounds using a Hurwicz criterion indexed by a parameter that governs the weight placed on the upper bound. We then use survey data from 2,050 Dutch adults to document that incomplete preferences over income, health, and social relations are empirically prevalent. Different resolutions of incompleteness have first-order distributive consequences. In particular, evaluating individuals at the lower bound systematically prioritizes those with the most incomplete preferences rather than those with the lowest outcomes: only 42 percent of individuals in the bottom decile under the lower-bound approach remain there under the upper-bound approach, and measured inequality varies by up to five Gini points. Measuring well-being under incomplete preferences therefore requires explicit normative choices about how interval-valued information is aggregated.

Keywords: Incomplete preferences; Interpersonal well-being comparisons; Equivalent income; Prioritarianism; Hurwicz criterion. **JEL-classification**: I30, I31, D63.

Acknowledgements: We are grateful to Marc Fleurbaey, Glenn Harrisson, Domenico Moramarco, Peter Wakker, and seminar participants in Antwerp, Montevideo, Bordeaux, Marseille, Paris, and Rotterdam for helpful comments. Financial support of the Research Foundation - Flanders (Excellence Of Science (EOS) grant G0G0318N) is gratefully acknowledged.

Corresponding authors: Santiago Burone (santiagogerman.buroneschaffner@uantwerpen.be).

1 Introduction

Economic analysis routinely relies on complete well-being rankings to identify the worst-off, measure inequality, and guide redistributive policy, particularly under prioritarian approaches that assign greater weight to individuals at the bottom of the well-being distribution (Adler and Norheim, 2022).[1] In many relevant settings, however, individual preferences are incomplete.[2] When outcomes are multidimensional—such as income, health, and social relations—individuals may be unable or unwilling to rank some alternatives, even in principle. When preference information is incomplete, can well-being be measured without taking a substantive normative stand on how unresolved comparisons are treated?

This paper shows that the answer is no. Incomplete preferences introduce an unavoidable normative choice in well-being measurement. When preference information is interval-valued rather than point-identified, any procedure that delivers policy-relevant complete rankings must resolve incomparability in a substantive way. Different resolutions of incompleteness, even when all respect individual preferences, can therefore lead to sharply different distributive judgments.

We establish this result by combining axiomatic analysis with empirical evidence. We first characterize a family of well-being measures in the spirit of Hurwicz (1951) in which individual well-being is evaluated as a weighted combination of the lower and upper bounds of the interval, with a parameter α reflecting the weight placed on the upper bound.[3] The parameter α captures how incomplete preferences are aggregated across individuals, with direct implications for identifying the worst-off and measuring inequality. The framework encompasses a spectrum from $\alpha = 0$ (lower bound only), $\alpha = 0.5$ (mid-point interval), to $\alpha = 1$ (upper bound only). Our goal is not to defend a particular choice for α, but to show that no neutral choice exists. In a prioritarian analysis, lower values prioritize individuals for whom low well-being cannot be ruled out—that is, those with low lower bounds; higher values prioritize individuals for whom high well-being can be ruled out—that is, those with low upper bounds. The former implements what Fleurbaey and Schokkaert (2013) call a "safety principle"; the latter reserves priority for those whose low well-being is precisely identified.

[1] The assumption of completeness underlies canonical contributions in social choice theory (Arrow, 1951), optimal tax theory (Mirrlees, 1971), demand-based welfare analysis and equivalent income measures (Deaton and Muellbauer, 1980), and recent frameworks for multidimensional welfare measurement (Fleurbaey and Maniquet, 2011).

[2] Behavioral economics (Bernheim and Rangel, 2009) and psychological research on judgment and decision-making (Lichtenstein and Slovic, 2006) have challenged the assumption of complete preferences. For a normative framework addressing incompleteness in multidimensional well-being comparisons, see Fleurbaey and Schokkaert (2013).

[3] Conceptually, this resembles the α-maxmin criteria in decision theory (Gilboa and Schmeidler, 1989), in which a weight interpolates between pessimistic and optimistic evaluations.

Using a survey of 2,050 Dutch adults, we elicit bounds on individuals' equivalent incomes–our measure of well-being–using an adaptive method that allows respondents to express incomparability directly—by selecting "Don't know"—rather than forcing a choice (see also Decancq and Nys, 2021). We find that 26 percent of respondents explicitly report being unable to rank their own life situation against alternatives. Incompleteness typically emerges early in the elicitation process, is associated with longer response times, and is systematically related to the respondents' actual health level. Together, these patterns indicate that incomplete preferences are quantitatively important and suggest genuine evaluative difficulty rather than inattention, satisficing behavior (Krosnick, 1991), or classical measurement error.[4]

Importantly, we show that the seemingly cautious lower-bound approach can backfire in a systematic and quantitatively important way. The manner in which incompleteness is resolved—the choice of α—fundamentally alters who counts as worst-off. Only 42 percent of individuals in the bottom decile under the lower-bound approach remain in the bottom decile under the upper-bound approach. At $\alpha = 0$, two thirds of the worst-off explicitly selected "Don't know" during elicitation and exhibit higher levels of income; they have low well-being primarily because their equivalent income intervals are wide, not because they have low income. In other words, the lower-bound approach systematically prioritizes individuals with the most incomplete preferences rather than those with the lowest measured outcomes—the groups prioritarians typically seek to protect. As greater weight is placed on upper bounds, the composition of the worst-off shifts toward individuals whose well-being is precisely measured. Evaluating at the midpoint of the interval ($\alpha = 0.5$) produces a compromise composition, blending individuals with wide intervals and those with precisely measured low well-being.

The choice of α also has substantial implications for measured inequality. Gini coefficients vary by up to five points depending on whether lower or upper bounds are used, with lower-bound evaluations producing systematically higher measured inequality in our data—reflecting that individuals with wide intervals are pulled down relative to those with precisely measured well-being.

Our results reveal a fundamental dilemma for prioritarian social evaluation. Lower-bound approaches effectively treat incompleteness as low well-being, while placing greater weight on upper bounds risks overlooking individuals whose incomplete preferences leave open the possibility that they are badly off. If priority is assigned solely on the basis of preference incompleteness, individuals with precisely identified low well-being are displaced from priority status. Whether this is acceptable depends on whether prioritarian concern is understood as

[4] We remain agnostic about the precise source of incompleteness in this paper. It may reflect cognitive difficulty with unfamiliar trade-offs, an incomplete process of preference construction, or fundamental incommensurability across life dimensions.

responding to low well-being that is precisely identified or as guarding against the possibility of being badly off. There is no normatively neutral choice: how incomplete preferences are aggregated directly shapes distributional outcomes, from the identification of the worst-off to measured inequality.

The paper's central contribution is to show that incomplete preferences make well-being measurement unavoidably normative, with potentially large distributive consequences. We formalize this insight by providing an axiomatic characterization of well-being measures under incomplete preferences. We show that any measure satisfying four natural axioms must resolve incomparability by aggregating equivalent income bounds using a Hurwicz-type criterion, in which the parameter α reflects the weight placed on the upper bound of the interval. Since no further principle within this framework determines α, the choice is irreducibly normative. Using original survey data, we then show that preference incompleteness is empirically prevalent and systematically related to individual characteristics, making this normative choice quantitatively relevant. Finally, we demonstrate that alternative normative choices for aggregating incomplete preferences have major implications for distributive judgments, shaping both who counts as worst-off and measured inequality. These insights extend beyond well-being measurement to any setting in which individual welfare or advantage is interval-valued rather than point-valued.

This paper contributes to three literatures. First, we build on work on equivalent incomes and multidimensional well-being measurement (Fleurbaey and Blanchet, 2013; Decancq et al., 2015a), exploring how incomplete preferences affect the empirical implementation of equivalent income measures and the aggregation of well-being. Second, we connect to the literature on incomplete preferences in decision theory (Bewley, 2002) and experimental work on preference imprecision and incompleteness (Butler and Loomes, 2007; Nielsen and Rigotti, 2023), providing new empirical evidence on their prevalence in welfare comparisons. Third, we contribute to normative analyses of prioritarianism (Parfit, 1984; Adler and Norheim, 2022) by showing that seemingly cautious aggregation rules can produce paradoxical outcomes when preferences are incomplete. Our findings reveal a tension in the "safety principle" of Fleurbaey and Schokkaert (2013) when incompleteness is widespread.

The remainder of the paper proceeds as follows. Section 2 develops the theoretical framework and presents the axiomatic characterization of well-being measures under incomplete preferences. Section 3 describes the elicitation method. Section 4 introduces the data from the Dutch LISS panel. Section 5 presents the empirical results. Section 6 concludes.

2 Theoretical Framework

2.1 Equivalent Income under Incomplete Preferences

A life situation $\ell = (y, z)$ consists of income $y \in \mathbb{R}_{++}$ and a k-dimensional non-income bundle $z \in Z$, where Z is a compact subset of $\mathbb{R}^k$. Let $\mathcal{L}$ denote the set of all such life situations. In our application, $k = 2$ and $z = (h, s)$ captures health and social relations.

Individuals have preferences P over life situations, where $\ell\, P\, \ell'$ means that ℓ is strictly preferred to ℓ'. For any ℓ and P, the upper and lower contour sets are

$$UC(\ell, P) = \{\ell' \mid \ell'\, P\, \ell\} \quad \text{and} \quad LC(\ell, P) = \{\ell' \mid \ell\, P\, \ell'\}.$$

We assume that P is asymmetric, transitive, monotone in income, and continuous, but not necessarily complete.[5] Let $\mathcal{P}$ denote the set of preference relations that satisfy these properties.

A *well-being measure* w assigns well-being levels to life situation-preference pairs (ℓ, P). A well-being measure is preference-based if it respects P: whenever $\ell' \in UC(\ell, P)$, $w(\ell', P) > w(\ell, P)$.

We use equivalent incomes to construct such a preference-based (Fleurbaey and Blanchet, 2013; Fleurbaey and Schokkaert, 2013; Decancq et al., 2015b). Fix a reference non-income bundle $\overline{z} \in Z$.[6] The equivalent income of ℓ, given preferences P, consists of all income levels at $\overline{z}$ that are neither strictly preferred nor strictly worse than ℓ. When preferences are complete, this is a single value. When preferences are incomplete, it is an interval (Fleurbaey and Schokkaert, 2013).[7] The *equivalent income interval* is bounded by

$$\begin{aligned} \underline{y}^\star(\ell, P) &= \sup\{y \mid (y, \overline{z}) \in LC(\ell, P)\}, \\ \overline{y}^\star(\ell, P) &= \inf\{y \mid (y, \overline{z}) \in UC(\ell, P)\}. \end{aligned} \tag{1}$$

These bounds define the range of incomparability: any $(y, \overline{z})$ with $\underline{y}^\star(\ell, P) < y < \overline{y}^\star(\ell, P)$ is incomparable to ℓ, i.e., it is neither strictly preferred nor strictly worse than ℓ. The width of this

[5] Preference relation P is (i) *asymmetric*: $\ell\, P\, \ell'$ implies not $\ell'\, P\, \ell$; (ii) *transitive*: if $\ell\, P\, \ell'$ and $\ell'\, P\, \ell''$, then $\ell\, P\, \ell''$; (iii) *monotone in income*: for fixed z, if $y > y'$, then $(y, z)\, P\, (y', z)$; (iv) *continuous*: $UC(\ell, P)$ and $LC(\ell, P)$ are open; (v) *complete*: for all ℓ, ℓ', either $\ell' \in UC(\ell, P)$ or $\ell' \in LC(\ell, P)$.

[6] The reference bundle is chosen such that for any ℓ and P, there exist income levels y, y' with $(y, \overline{z}) \in UC(\ell, P)$ and $(y', \overline{z}) \in LC(\ell, P)$. The choice of $\overline{z}$ reflects a normative judgment (Fleurbaey and Blanchet, 2013): it identifies the conditions under which income alone suffices for well-being comparisons. In our application, we select perfect health and social relations as the reference.

[7] By monotonicity, continuity, and our choice of reference bundle, the bounds of the interval exist, are finite, and vary continuously with ℓ. In the following, we suppress the dependence on $\overline{z}$ in the notation.

interval,

$$\Delta y^{\star}(\ell, P) = \overline{y}^{\star}(\ell, P) - \underline{y}^{\star}(\ell, P),$$

measures the local incompleteness of P at ℓ (see Karni and Vierø (2023) for an analogous measure in the context of choice under uncertainty). Figure 1 illustrates the upper and lower contour sets and the incomparable set (shaded) for the case of a single non-income dimension ($k = 1$). It also shows the equivalent income interval and its bounds.

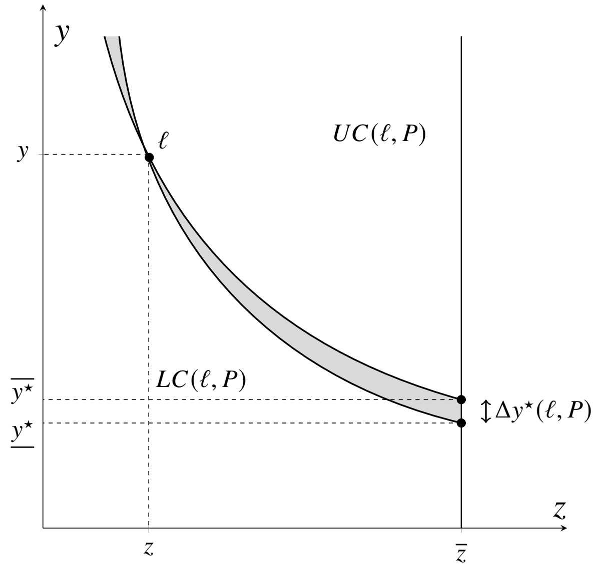

Figure 1 Equivalent Income Interval

Notes: The figure depicts the upper and lower contour sets, $UC(\ell, P)$ and $LC(\ell, P)$, together with the shaded incomparable set. In addition, it shows the equivalent income interval and its corresponding bounds $\underline{y}^{\star} = \underline{y}^{\star}(\ell, P)$ and $\overline{y}^{\star} = \overline{y}^{\star}(\ell, P)$.

A preference relation $P' \in \mathcal{P}$ is a *refinement* of $P \in \mathcal{P}$ at ℓ if it is more complete at ℓ in the sense that:

$$UC(\ell, P) \subseteq UC(\ell, P') \quad \text{and} \quad LC(\ell, P) \subseteq LC(\ell, P'),$$

with at least one strict inclusion. A refinement narrows the interval of incomparability and lowers the local measure of incompleteness. The impact of a refinement on the local incompleteness measure admits a natural decomposition into an upper bound effect and a lower bound effect. For any refinement P' of P at ℓ:

$$\Delta y^{\star}(\ell, P) = \Delta y^{\star}(\ell, P') + \underbrace{\left(\overline{y}^{\star}(\ell, P) - \overline{y}^{\star}(\ell, P')\right)}_{\text{Upper bound effect}} + \underbrace{\left(\underline{y}^{\star}(\ell, P') - \underline{y}^{\star}(\ell, P)\right)}_{\text{Lower bound effect}}. \qquad (2)$$

Although both upper and lower bound effects reduce incomparability, they have an opposite impact on preference-based well-being, given that preferences are monotone in income: expanding the upper contour set (weakly) lowers well-being, while expanding the lower contour set (weakly) raises it. The next section discusses how these effects can be aggregated.

2.2 Hurwicz Equivalent Income Measure

Many policy applications require complete well-being rankings rather than incomplete ones. Following Hurwicz (1951), we characterize a well-being measure w that aggregates the bounds $[\underline{y}^\star(\ell, P), \overline{y}^\star(\ell, P)]$ into a complete ordering (see also Arrow and Hurwicz, 1977). We show that a natural set of four axioms uniquely pins down the Hurwicz equivalent income well-being measure as a convex combination of the lower and upper bounds.

Axiom 1 (Bound Invariance)**.** *For all $\ell, \ell' \in \mathcal{L}$ and $P, P' \in \mathcal{P}$, if $\underline{y}^\star(\ell, P) = \underline{y}^\star(\ell', P')$ and $\overline{y}^\star(\ell, P) = \overline{y}^\star(\ell', P')$, then $w(\ell, P) = w(\ell', P')$.*

The axiom Bound Invariance states that individuals with the same equivalent income interval have the same well-being.

Axiom 2 (Monotonicity in Bounds)**.** *For all $\ell, \ell' \in \mathcal{L}$ and $P, P' \in \mathcal{P}$, if $\underline{y}^\star(\ell, P) \geq \underline{y}^\star(\ell', P')$ and $\overline{y}^\star(\ell, P) \geq \overline{y}^\star(\ell', P')$, then $w(\ell, P) \geq w(\ell', P')$.*

This axiom establishes a weak partial ordering: when both bounds are weakly higher, well-being is weakly higher.[8]

Axiom 3 (α-balanced Invariance)**.** *For all $\ell, \ell' \in \mathcal{L}$ and $P, P' \in \mathcal{P}$, let $\alpha \in [0, 1]$. If $\overline{y}^\star(\ell, P) > \overline{y}^\star(\ell', P')$, $\underline{y}^\star(\ell', P') > \underline{y}^\star(\ell, P)$, and*

$$\alpha(\overline{y}^\star(\ell, P) - \overline{y}^\star(\ell', P')) = (1 - \alpha)(\underline{y}^\star(\ell', P') - \underline{y}^\star(\ell, P)),$$

then $w(\ell, P) = w(\ell', P')$.

This axiom determines how the opposing bound effects from equation (2) are aggregated. Well-being is unchanged when the upper bound effect (weighted by α) exactly offsets the lower bound effect (weighted by $1 - \alpha$). The parameter α therefore captures the relative weight placed on upper bound versus lower bound adjustments: higher values place more weight on what is strictly preferred, while lower values place more weight on what is strictly worse.

[8] Fleurbaey and Schokkaert (2013) derive Monotonicity in Bounds as a consequence of their "super safety"-principle.

Crucially, the axiom does not introduce a new normative judgment but makes an unavoidable one explicit. Any procedure that transforms interval-valued preference information into complete, policy-relevant well-being rankings must specify how upper and lower bounds are traded off. Axiom 3 imposes two deliberate and minimal restrictions: aggregation is linear, and the weight α is constant across individuals and situations. These restrictions rule out level-dependent or reference-dependent treatments of incompleteness, while delivering the simplest structure that nests common practices and allows transparent sensitivity analysis. Rather than embedding normative choices implicitly, the axiom isolates them in a single, interpretable parameter.

Axiom 4 (Normalization)**.** *For all $\ell \in \mathcal{L}$ and $P \in \mathcal{P}$, if $\underline{y}^\star(\ell, P) = \overline{y}^\star(\ell, P) = y$, then $w(\ell, P) = y$.*

This axiom fixes the cardinal well-being level when preferences are complete. It establishes that well-being equals equivalent income in the standard case where the interval is degenerate.

The following theorem provides a complete characterization of the family of well-being measures that satisfy the four axioms. No other functional form satisfies them all.

Theorem 1. *A well-being measure $w : \mathcal{L} \times \mathcal{P} \to \mathbb{R}_+$ satisfies Bound Invariance, Monotonicity in Bounds, α-balanced Invariance (for some $\alpha \in [0, 1]$), and Normalization if and only if, for all $\ell \in \mathcal{L}$ and $P \in \mathcal{P}$,*

$$w(\ell, P) = (1 - \alpha)\, \underline{y}^\star(\ell, P) + \alpha\, \overline{y}^\star(\ell, P).$$

Proof. See Appendix A.

The theorem characterizes a family of well-being measures indexed by $\alpha \in [0, 1]$. We call α the incompleteness attitude and denote by $w_\alpha(\ell, P)$ the Hurwicz equivalent income corresponding to α, defined as:

$$w_\alpha(\ell, P) = (1 - \alpha)\, \underline{y}^\star(\ell, P) + \alpha\, \overline{y}^\star(\ell, P). \tag{3}$$

The equivalent income bounds $\left(\underline{y}^\star(\ell, P), \overline{y}^\star(\ell, P)\right)$ constitute sufficient statistics for well-being within this family of measures: all information relevant for well-being evaluation is given by the bounds, with α determining how they are weighted.[9] A multidimensional evaluation problem over life situations and incomplete preferences thus reduces to a two-dimensional representation in the space of bounds.

The incompleteness attitude α admits a direct prioritarian interpretation. Lower values of α prioritize individuals for whom low well-being cannot be ruled out—those whose interval

[9] The life situation ℓ and preference relation P affect well-being only through their effect on the bounds, with no additional direct effect. For notational convenience, we henceforth suppress the dependence of the lower and upper bounds on (ℓ, P).

extends far downward, regardless of its upper bound. Higher values prioritize individuals for whom high well-being can be ruled out—those whose interval is bounded far downward, so that even its upper bound is low. Intermediate values interpolate between these positions. Choosing $\alpha = 0$ treats incompleteness itself as grounds for priority: individuals whose preferences leave open very low equivalent incomes receive priority even if their intervals also extend upward. This implements a version of the "safety principle" (Fleurbaey and Schokkaert, 2013), according to which one should avoid ranking an individual as better off when the available information is compatible with them being worse off. At the other extreme, choosing $\alpha = 1$ assigns priority only on the basis of precisely identified low well-being, ignoring incomparability altogether. Neither position is normatively compelled; the choice of α reflects a substantive judgment about how incomplete preferences should enter distributive evaluation.

Having characterized how to measure the well-being of a single individual, we now turn to the question of how to compare well-being across individuals with potentially different degrees of preference incompleteness. When comparing well-being across individuals, a natural question arises: which value of α should be used? Different values may yield different rankings. When no single value is compelling, we can consider rankings that hold across a range of α values. We formalize three criteria that differ in how much agreement about α they require.

Definition 1 (Non-overlap criterion)**.** *For all $\ell, \ell' \in \mathcal{L}$ and $P, P' \in \mathcal{P}$, (ℓ, P) is better off than (ℓ', P') whenever $w_\alpha(\ell, P) > w_{\alpha'}(\ell', P')$ for all $\alpha, \alpha' \in [0, 1]$.*

The non-overlap criterion requires unanimity: (ℓ, P) ranks above (ℓ', P') for all possible values of α chosen for each individual. This holds if and only if $\underline{y}^\star(\ell, P) > \overline{y}^\star(\ell', P')$, that is, when the equivalent income intervals do not overlap.

Definition 2 (Dominance criterion)**.** *For all $\ell, \ell' \in \mathcal{L}$ and $P, P' \in \mathcal{P}$, (ℓ, P) is better off than (ℓ', P') whenever $w_\alpha(\ell, P) > w_\alpha(\ell', P')$ for all $\alpha \in [0, 1]$.*

The dominance criterion requires that (ℓ, P) rank above (ℓ', P') when both individuals are evaluated using the same incompleteness attitude α, for all $\alpha \in [0, 1]$. Given the linear form of w_α, this holds if and only if both bounds are weakly higher: $\underline{y}^\star(\ell, P) \geq \underline{y}^\star(\ell', P')$ and $\overline{y}^\star(\ell, P) \geq \overline{y}^\star(\ell', P')$, with at least one strict inequality. This criterion implements the partial ordering implied by Monotonicity in Bounds (Axiom 2) for interpersonal comparisons.

Definition 3 (Hurwicz criterion)**.** *For all $\ell, \ell' \in \mathcal{L}$ and $P, P' \in \mathcal{P}$, and for a fixed $\alpha \in [0, 1]$, (ℓ, P) is better off than (ℓ', P') whenever $w_\alpha(\ell, P) > w_\alpha(\ell', P')$.*

The three criteria are nested in terms of their strength.[10] The non-overlap criterion is most demanding, requiring that (ℓ, P) rank above (ℓ', P') for all possible combinations of α and

[10] Each approach defines an asymmetric and transitive relation for well-being comparisons.

α'. The dominance criterion relaxes this by requiring only that the same α be used for both individuals. The complete ranking with fixed α is least demanding, but the resulting ranking depends on the choice of α. Formally, if (ℓ, P) satisfies the non-overlap criterion relative to (ℓ', P'), it also satisfies the dominance criterion, which in turn implies the Hurwicz criterion. For any fixed α, the Hurwicz criterion provides a complete well-being ranking. The non-overlap and dominance criteria are incomplete: not all pairs of individuals can be ranked. The share of pairs that can be ranked by each criterion is an empirical question examined in Section 5.

3 Preference Elicitation Method

To elicit the bounds of the equivalent income interval $[\underline{y}^\star, \overline{y}^\star]$ defined in Equation (1), we use the Adaptive Bisectional Dichotomous Choice (ABDC) method proposed by Decancq and Nys (2021).[11] Respondents are presented with a sequence of dichotomous choices between their own life situation ℓ and alternative life situations $\ell'_r = (y^\star_r, \overline{z})$, where the non-income bundle is fixed at the reference level $\overline{z}$ and income $y^\star_r$ varies across rounds $r = 1, 2, \ldots, R$.

At each round, the income level $y^\star_r$ is updated adaptively and set at the midpoint of the current equivalent income interval implied by the respondent's previous choices:

$$y^\star_r = \frac{\underline{y}^\star_{r-1} + \overline{y}^\star_{r-1}}{2}.$$

The algorithm initializes with $\underline{y}^\star_0 = 0$ and $\overline{y}^\star_0 = y$. Depending on the respondent's choice, the midpoint replaces one of the bounds:

$$\begin{aligned} \underline{y}^\star_r &= y^\star_r \quad \text{if } \ell'_r = (y^\star_r, \overline{z}) \in LC(\ell, P), \\ \overline{y}^\star_r &= y^\star_r \quad \text{if } \ell'_r = (y^\star_r, \overline{z}) \in UC(\ell, P). \end{aligned}$$

The ABDC elicitation exploits monotonicity in income and transitivity of preferences in $\mathcal{P}$ to ensure that the equivalent income interval $[\underline{y}^\star_r, \overline{y}^\star_r]$ is well defined at each round r for the reference non-income bundle $\overline{z}$. Figure 2 illustrates the procedure.[12] At each round r, the algorithm refines the preference relation and halves the local incompleteness measure $\Delta y^\star_r = \overline{y}^\star_r - \underline{y}^\star_r$.

A distinctive feature of our elicitation strategy is that respondents are not forced to make complete comparisons. In each dichotomous choice, respondents may indicate that they are unable to rank their own life situation relative to the alternative. Such "Don't know" responses

[11] For a discussion of the ABDC method, see Da Costa et al. (Forthcoming). Cabeza Martínez and Decancq (2025) use the method to elicit social preferences and measure altruism.

[12] Appendix B provides pseudocode for the algorithm and a screenshot of a dichotomous choice.

are interpreted as revealing incomparability (i.e., $\ell'_r \notin LC(\ell, P)$ and $\ell'_r \notin UC(\ell, P)$) and thus incompleteness of the underlying preference relation, rather than as missing or noisy data. This design choice aligns the elicitation procedure with our theoretical framework of incomplete preferences and avoids imposing artificial completeness through forced choices.[13]

Selecting "Don't know" terminates the algorithm.[14] The procedure also stops when the maximum number of rounds is reached (i.e., $r = R$, with $R = 10$ in our application) or when the desired precision is achieved ($\Delta y^\star_r \leq 1$ euro in our application).[15]

Compared to direct interval contingent valuation methods using show cards, the ABDC procedure elicits bounds through a sequence of simple binary choices, thereby reducing respondents' cognitive burden (Attema and Brouwer, 2013) and fatigue. The bisectional updating rule further ensures rapid (exponential) reduction of the elicitation incompleteness across rounds, faster than when lists of intervals are shown. A potential limitation is sensitivity to the income level shown in the first round, which may induce starting-point bias toward income levels near one-half of own income (Burone and Leitner, 2025).[16]

This elicitation procedure gives rise to two conceptually distinct sources of incompleteness, only one of which reflects genuine evaluative difficulty. Intrinsic incompleteness occurs when respondents select the "Don't know" option, revealing incomparability between ℓ and ℓ'_r at the moment of elicitation. Such incompleteness may reflect fundamental incommensurability across life dimensions, genuine uncertainty about one's own preferences, or an incomplete process of preference construction (Lichtenstein and Slovic, 2006). Elicitation incompleteness, by contrast, arises mechanically from the finite precision of the algorithm, that is, from $\Delta y^\star_r > 0$ after the last elicitation round. The latter decreases deterministically with each round of the ABDC procedure.

Importantly, our framework does not require taking a position on which of these sources is operative. Whether incompleteness reflects intrinsic evaluative difficulty or finite elicitation precision, the measurement problem is the same: the information available for assessing individual well-being is interval-valued rather than point-identified. Any policy application that requires complete interpersonal comparisons—such as identifying the worst-off or measuring inequality—must therefore aggregate these bounds in a principled way. The Hurwicz framework

---

[13] On the use of "Don't know" responses in the direct elicitation of incomplete preferences, see Nielsen and Rigotti (2023).

[14] See Algorithm 1 for pseudocode of the complete algorithm. It yields conservative estimates in the sense that the elicited interval weakly contains the true equivalent income interval implied by ℓ and P.

[15] After r rounds, $\Delta y^\star_r$ equals $1/2^r$ of the initial income level y. In our application, the average income is approximately 1,800 euros, so that $\Delta y^\star_r \leq 1$ euro would typically be achieved when $r = 11$.

[16] Burone and Leitner (2025) provide experimental evidence that the initial alternative life situation ℓ'_1 in the ABDC method can anchor subsequent responses, implying that our equivalent income measures may be biased toward half of respondents' income.

does not eliminate this choice; it makes it explicit and isolates its normative content in a single parameter—the incompleteness attitude α.

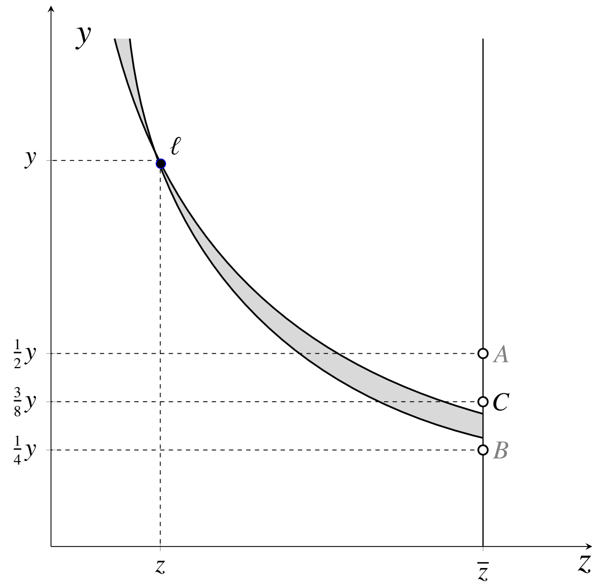

Figure 2 Illustration of the ABDC procedure

Notes: The figure illustrates the Adaptive Bisectional Dichotomous Choice (ABDC) procedure. In the first round, the respondent compares the life situation ℓ with A. Since A is preferred, the respondent then compares ℓ with B in the next round, followed by a comparison with C, and so on.

4 Data

We use data from a dedicated module we fielded in February 2022 in the LISS (Longitudinal Internet Studies for the Social Sciences), a probability-based internet panel in the Netherlands. The panel consists of Dutch-speaking households randomly drawn from the population register. Fieldwork took place during the later stages of the COVID-19 pandemic in the Netherlands. While the dimensions of income, health, and social relations may have been particularly salient during this period, they are standard in the equivalent income literature (see Da Costa et al., Forthcoming, for a recent survey).

Invitations were sent to 3,269 randomly selected panel members aged 25–75, of whom 2,543 participated (77.8 percent response rate). After excluding respondents in an experimental subsample, those with incomplete responses, zero income, inconsistent preferences, or flagged comments, the full sample consists of 2,186 respondents. Following Stantcheva (2023), we further exclude 136 speeders, yielding a baseline sample of 2,050.[17] To address selection on observables into the baseline sample, we apply inverse-probability weights.[18]

The questionnaire began with an introductory screen explaining the study's purpose, emphasizing the voluntary nature of participation, and assuring respondents that their answers would remain anonymous and confidential.[19] Respondents were then shown pre-computed measures of their income, health, and social relations—derived from earlier LISS survey waves—and asked to confirm or adjust these values. Income corresponds to monthly net per capita household income (in euros). Health and social relations were measured using composite indices constructed from previously provided information.[20] Most respondents accepted the pre-loaded values, and adjustments were generally small.[21]

After an example choice task, respondents completed a training question in which they chose between their current life ℓ and an alternative $\ell_0 = (y, \bar{z})$ with their actual income and perfect health and social relations. This familiarized respondents with the choice format and allowed a consistency check.[22] The main module then consisted of the ABDC questions used to elicit equivalent income bounds and identify preference incompleteness; average duration was approximately eight minutes. Respondents concluded with a brief questionnaire evaluation.

Table 1 reports summary statistics. Average monthly net per capita household income is approximately €1,800 with substantial dispersion. Health and social relations indices range from 0 to 100, with mean values around 71. The average respondent is 55 years old, and the sample is broadly balanced by gender. A majority is married; unemployment is low, and roughly 25 percent are pensioners. About 16 percent hold a university degree, and approximately 17 percent report a non-Dutch background.

[17] Respondents were classified as speeders if they spent insufficient time on survey screens; see Appendix D.1 for details.

[18] Older respondents, pensioners, and men were more likely to remain in the baseline sample. See Appendix D.1 for details; Appendices E.1 and E.2 report sensitivity analyses using the unweighted baseline sample and the full sample including speeders. Results are robust.

[19] LISS panel members consent to their data being used for scientific research. Formal ethical approval is not required under Dutch law for questionnaire research with adults; see lissdata.nl/ethics.

[20] Income was taken from the LISS Income Core Module. Health and social relations indices were constructed using information from the LISS "Health" and "Social Relations and Leisure" Core Modules. See Appendix D.2 for details.

[21] See Appendix C.1 for distributions of the original (pre-loaded) and adjusted well-being dimensions as well summary statistics of the adjustments.

[22] Comparing responses to the training question and the first elicitation round allowed us to check for transitivity and monotonicity violations; see Appendix D.1.

Table 1 Summary Statistics

Variable	Mean	Min	Max	SD	*N*
Income	1,798	188	73,708	2,103	2,050
Health index	71	0	100	18	2,050
Social relations index	71	0	100	16	2,050
Age	55	25	76	14	2,050
Female (%)	55	0	100	50	2,050
Married (%)	56	0	100	50	2,050
Unemployed (%)	2	0	100	15	2,050
Pensioner (%)	25	0	100	44	2,050
University degree (%)	16	0	100	36	2,050
Non-Dutch background (%)	17	0	100	37	2,032

Notes: Weighted baseline sample.

5 Empirical Results

5.1 Prevalence of Incomplete Preferences

Twenty-six percent of respondents selected "Don't know" at least once during elicitation. Respondents could always choose one of the two life situations, so this response reflects an active expression of incomparability rather than a forced default. The cumulative share rises rapidly in early elicitation rounds–reaching 18 percent by round 4 and 24 percent by round 7–before plateauing (Figure 3), suggesting that incomparability emerges early for a subset of respondents.

Response times also suggest that incompleteness reflects deliberation rather than inattention: respondents take consistently longer to respond in rounds that culminate in "Don't know" (Figure 12), suggesting greater difficulty in evaluating the trade-offs involved.

As the algorithm progresses, elicitation incompleteness declines mechanically for respondents who continue to make choices without selecting "Don't know." After ten rounds, the distribution of equivalent income interval widths $\Delta y^\star$ is right-skewed, with median width €2, 75th percentile about €10, and a long right tail extending beyond €2300 (Figure 13). Respondents who selected "Don't know" populate this upper tail, while the others have small interval widths.

5.2 Determinants of Incompleteness

To examine the determinants of incompleteness, we estimate two complementary models. First, we regress the probability of selecting "Don't know" on income, indices of health and social relations, demographic characteristics (age, gender, education), Big Five personality traits, and

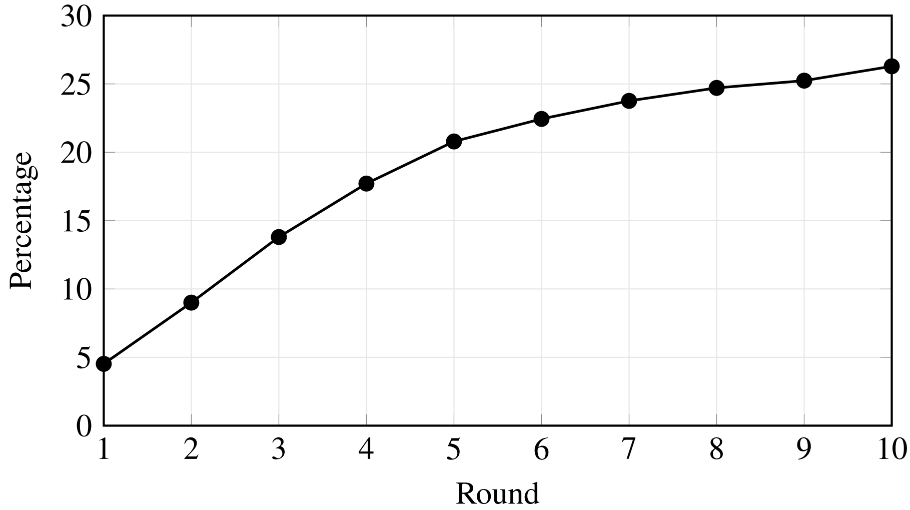

Figure 3 Cumulative Prevalence of "Don't know" Responses by Elicitation Round
Notes: Weighted baseline sample (N = 2,050). See Appendices E.1–E.2 for robustness.

survey evaluation measures, using a probit specification with robust standard errors. Second, we regress the local incompleteness measure–the equivalent income interval width– $\Delta y^\star$ on the same set of covariates using OLS.[23] Results are presented in Table 2.[24]

The estimates reveal several salient patterns. First, the health index is a predictor of incompleteness: a one-point increase in the health index is associated with a lower probability of selecting "Don't know" ($p < 0.05$) and a smaller final interval width ($p < 0.1$). Higher health levels imply a smaller distance to the reference value of perfect health, making the hypothetical life situations considered in the elicitation less distant from respondents' actual life situation and therefore easier to evaluate. Second, age is negatively associated with incompleteness ($p < 0.01$), whereas holding a university degree is positively associated with selecting "Don't know" ($p < 0.1$), consistent with greater evaluative caution or deliberation rather than lower comprehension.[25] Among the Big Five personality traits, only conscientiousness shows a marginally significant negative association with incompleteness ($p < 0.01$), while the other traits are not significantly

[23] The incompleteness measure $\Delta y^\star$ is mechanically larger for respondents who select "Don't know" early in the elicitation. The OLS specification should therefore be interpreted as examining variation in the degree of incompleteness among those who exhibit incomplete preferences.

[24] Appendix C.8 reports alternative specifications; findings are robust.

[25] Women are less likely to select "Don't know" overall ($p < 0.05$), but exhibit slightly higher rates in the earliest rounds (5.4 percent versus 4.2 percent in round 1; see Appendix C.2). This suggests initial caution followed by faster resolution, consistent with evidence viewing "Don't know" as caution rather than indecision (Bucher-Koenen et al., 2025; Harrison et al., Forthcoming).

Table 2 Determinants of Incompleteness

Dependent Variable	"Don't know"		$\Delta y^{\star}$	
Income (in logarithm)	0.107	(0.074)	0.011*	(0.008)
Health index	-0.006**	(0.002)	-0.001*	(0.000)
Social relations index	-0.001	(0.003)	-0.000	(0.000)
Female	-0.192**	(0.079)	-0.010	(0.008)
Older (above median)	-0.333***	(0.104)	-0.049***	(0.010)
University degree	0.208*	(0.109)	-0.002	(0.011)
Pensioner	-0.247**	(0.116)	-0.015	(0.013)
Works	-0.056	(0.104)	0.005	(0.010)
Openness	-0.024	(0.082)	0.011	(0.009)
Neuroticism	0.030	(0.063)	0.009	(0.006)
Conscientiousness	-0.178*	(0.100)	-0.007	(0.010)
Agreeableness	-0.097	(0.081)	-0.010	(0.008)
Extraversion	-0.041	(0.060)	-0.001	(0.006)
Did the questions make you think?	0.323***	(0.118)	0.011	(0.011)
Did you find the questions unclear?	0.487***	(0.139)	0.053***	(0.020)
Did you find the questions interesting?	-0.041	(0.144)	-0.015	(0.015)
Time last round	0.165***	(0.010)	0.028***	(0.001)
Constant	-1.285**	(0.618)	-0.089	(0.066)
Estimation method	Probit		OLS	
N	1,953		1,953	
(Pseudo) R^2	0.309		0.474	

Notes: Weighted baseline sample. See Appendices E.1–E.2 for robustness.
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

associated with either outcome.[26] Turning to the survey evaluation measures, respondents who reported that the tasks were "unclear" or "made me think" are significantly more likely to choose "Don't know" ($p < 0.01$) and to end the elicitation with wider intervals. Finally, spending more time looking at the screen before answering is strongly associated with a higher likelihood of selecting "Don't know" and with ending the elicitation with wider intervals ($p < 0.001$).

Taken together, these results provide evidence that incompleteness in preferences over multidimensional life situations is substantively meaningful rather than artifactual. That more than one-quarter of respondents actively select "Don't know" when comparing life situations indicates that incompleteness is a quantitatively important feature of individual preferences in this domain. Longer response times, the association with distance to the reference health level, the positive correlation with higher educational attainment, and respondents' own reports that the tasks "made me think" are all consistent with incompleteness reflecting the inherent difficulty of evaluating life-situation trade-offs, rather than being driven primarily by satisficing behavior, inattention, or lack of effort (Krosnick, 1991). Having established that incomplete preferences are

[26] This aligns with Sturgis and Brunton-Smith (2023), who find Conscientiousness negatively associated with "Don't know" responses in online surveys.

both prevalent and substantively meaningful, we now examine their implications for interpersonal comparisons and the identification of the worst-off.

5.3 Well-being Comparisons under Incompleteness

Incomplete individual preferences have direct implications for interpersonal comparisons. We examine how the three criteria defined in Section 2—non-overlap, dominance, and Hurwicz for a given α—trade off robustness to normative disagreement against the ability to rank pairs.[27]

Figure 4 displays the share of pairs ranked by each criterion after each ABDC round. The non-overlap criterion shows substantial incompleteness in early rounds but improves rapidly: by round 5, approximately 86 percent of pairs can be ranked, consistent with Decancq and Nys (2021). By round 10, this reaches 88 percent. The speed of convergence is informative: a small number of elicitation rounds suffices to rank most pairs.

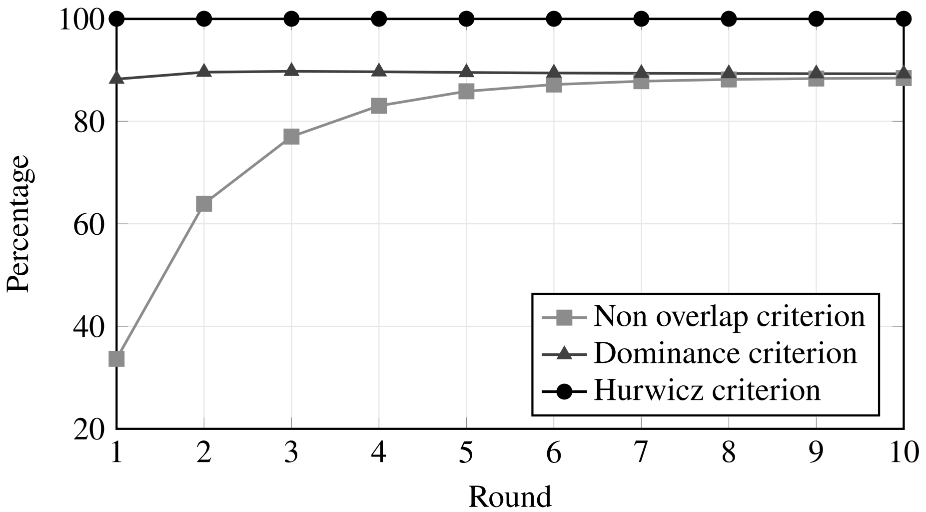

Figure 4 Share of pairs ranked after each elicitation round, by criterion
Notes: Weighted baseline sample (N = 2,050). See Appendices E.1–E.2 for robustness.

The dominance criterion ranks more pairs initially but stabilizes earlier. From round 1, the share plateaus at approximately 88 percent. Pairs that cannot be ranked are those whose ordering reverses between lower and upper bounds.[28] Such reversals are more likely when at least one individual selected “Don’t know.” Unlike non-overlap incompleteness, which diminishes with

[27] With 2,050 respondents, there are 2,100,225 unique pairs. Approximately 47 percent include at least one respondent who selected “Don’t know”; see Appendix C.5.

[28] Appendix C.6 shows the distribution of $\alpha^\star$ values at which rankings reverse. Most reversals occur at low α.

additional elicitation rounds, dominance incompleteness reflects irreducible normative ambiguity over how lower and upper bounds should be aggregated. Additional elicitation rounds cannot resolve it.

Unlike the non-overlap and dominance criteria, the Hurwicz criterion for a given α provides a complete ordering of all individuals. This completeness comes at the price of selecting a value for the incompleteness attitude α. Figure 5 shows how the distribution of Hurwicz equivalent incomes evolves as α varies from 0 to 1. For respondents who never selected “Don’t know,” the impact of α is negligible given their narrow equivalent income levels after 10 rounds. Respondents who selected “Don’t know” exhibit substantial sensitivity: at $\alpha = 0$, a large mass appears in the lower tail, but these individuals move progressively upward as α increases toward 1. This raises a fundamental question for policymakers who require complete rankings: which α should be used to identify the worst-off?

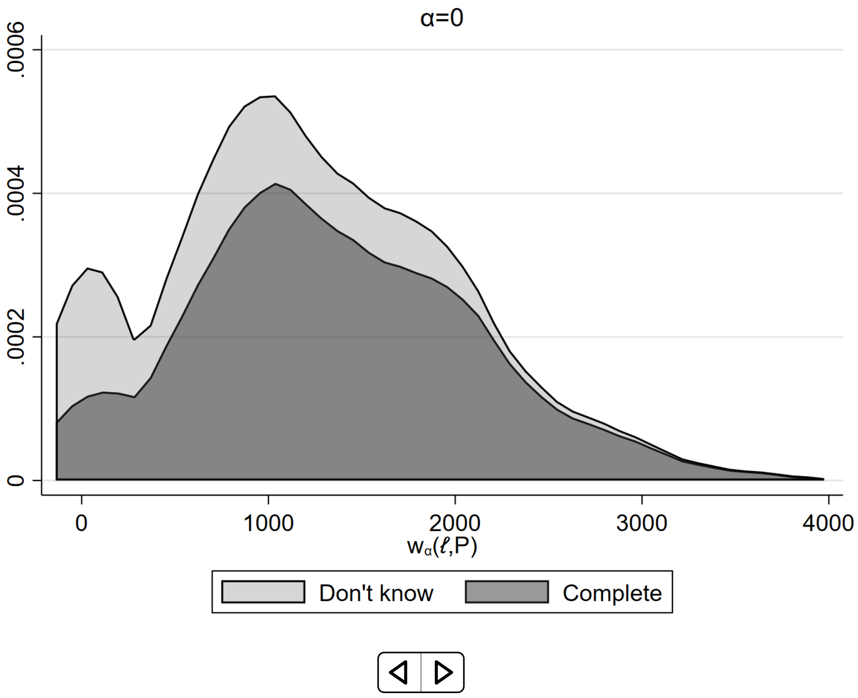

Figure 5 Hurwicz equivalent incomes, across α values
Notes: Epanechnikov kernel density function of Hurwicz equivalent incomes across values of the α parameter (Bandwidth: 138.5). The figure is animated; click the play button to observe how the distribution evolves as α increases. A static version is provided in Appendix C.9. Weighted baseline sample (N = 2,050). See Appendices E.1–E.2 for robustness.

5.4 Implications for Prioritarians

For prioritarians seeking to identify the worst-off, the choice of α governs a fundamental trade-off in how priority is assigned. Lower values of α prioritize all individuals for whom low well-being cannot be ruled out—that is, those with low lower bounds $\underline{y}^{\star}$. Higher values prioritize individuals for whom high well-being can be ruled out—those with low upper bounds $\overline{y}^{\star}$. As discussed in Section 2, the choice is a normative one; we now examine how this choice affects distributional judgments in practice.

Table 3 reports characteristics of individuals classified as worst-off (bottom decile) under alternative values of α. The composition of this group varies sharply with α: the overlap between the bottom decile at $\alpha = 0$ and $\alpha = 1$ is only 45 percent.[29] This limited overlap reveals that the choice of α fundamentally alters the identity of those considered worst-off.

Table 3 Characteristics of the Worst-Off

	$\alpha = 0$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$
Income	1606	1237	1111	1097	1075
Health index	57	58	58	58	59
Social relations index	64	62	63	63	64
Age	57	54	52	53	52
Female (%)	60.8	61.0	60.9	62.8	62.2
Married (%)	60.4	65.5	66.1	65.8	64.3
Unemployed (%)	3.6	5.5	6.3	6.7	6.2
Pensioner (%)	27.5	16.4	13.3	13.3	13.6
University degree (%)	10.8	9.6	10.0	10.3	10.3
Non-Dutch background (%)	24.6	28.3	26.5	26.5	26.5
Don't know (%)	64.8	45.7	31.6	26.5	20.3
Rounds	4.5	6.3	7.6	8.0	8.5
N	205	205	205	205	205

Notes: Each column reports averages for the bottom decile under the Hurwicz equivalent-income measure for a given α. Weighted baseline sample. See Appendices E.1–E.2 for robustness.

At $\alpha = 0$, which implements the seemingly cautious choice of placing full weight on lower bounds $\underline{y}^{\star}$, the worst-off group is dominated by respondents with incomplete preferences: 65 percent selected "Don't know," and the average number of ABDC rounds completed is only 4.5. Strikingly, these individuals do not exhibit particularly low income compared to the total baseline sample. While they have lower health and social relations, the characteristics of the worst-off are not different from the total sample (see Table 1). They appear at the bottom of the distribution

[29] See Appendix C.7 for the full transition matrix.

primarily because their equivalent income intervals are wide, rather than because their observed outcomes are especially low.

As α increases, the composition of the worst-off shifts dramatically. The prevalence of incomplete preferences declines monotonically—from 65 percent at $\alpha = 0$ to 20 percent at $\alpha = 1$—while those with lower income and weaker labor market attachment are found more among the worst-off. At the same time, the number of completed ABDC rounds rises, indicating more precisely identified equivalent income bounds. Higher values of α thus place greater weight on low well-being that is precisely identified rather than on incompleteness itself.

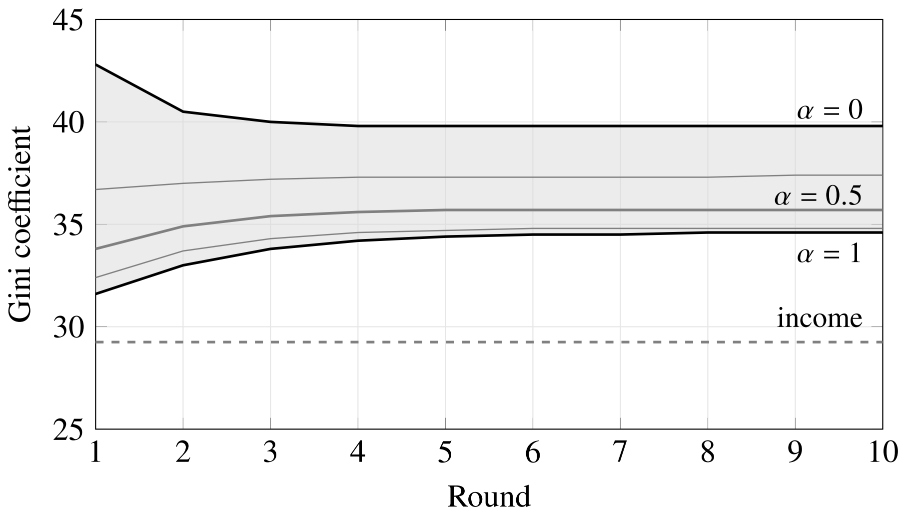

Figure 6 Bounds on Well-Being Inequality by Elicitation Round
Notes: Gini coefficient of the Hurwicz equivalent incomes across α values, after each elicitation round of the ABDC procedure. Dotted line indicates the Gini coefficient of the income dimension. Weighted baseline sample (N = 2,050). See Appendices E.1–E.2 for robustness.

The choice of α also affects measured inequality. Figure 6 illustrates bounds on well-being inequality across elicitation rounds, where the shaded area captures the range of Gini coefficients implied by $\alpha \in [0, 1]$. For reference, the dotted line shows the Gini coefficient of the income dimension, which equals 29.3. Well-being inequality, as measured by equivalent income, is considerably higher. As elicitation progresses, equivalent income intervals narrow and the bounds on the Gini coefficient narrow as well. After 10 elicitation rounds, the difference between $\alpha = 0$ and $\alpha = 1$ remains sizeable: about five Gini points.[30] Lower α produces higher measured inequality by placing full weight on lower bounds, whereas higher α compresses the distribution

[30] As shown in Table 8 in Appendix C.10, the gap between the upper and lower bounds widens substantially for S-Gini indices that assign more weight to the lower tail of the distribution (higher δ) compared to the standard Gini coefficient ($\delta = 2$).

(see also Figure 5). Even moderate intermediate weights ($\alpha = 0.25$ or 0.5) significantly reduce inequality relative to the lower bound ($\alpha = 0$). These patterns demonstrate that the choice of α has first-order implications for prioritarian judgments about both the identification of the worst-off and the extent of well-being inequality.

This trade-off poses a dilemma for prioritarians. While the cautious approach ($\alpha = 0$) embodies epistemic humility, it directs priority toward individuals with incomplete preferences rather than those with the lowest observed outcomes. The seemingly pro-poor lower-bound criterion thus conflates two distinct phenomena: genuinely low well-being and wide incomparability. Conversely, outcome-focused criteria (high α) risk disregarding the possibility that wide equivalent income intervals reflect genuine difficulty in evaluating severely adverse circumstances. Our evidence suggests that preference incompleteness is, at least in part, intrinsic: it emerges early in the elicitation, is associated with longer response times, and correlates systematically with distance from reference health levels—patterns that argue against mere satisficing behavior or measurement error.

6 Conclusion

This paper has examined the prevalence and consequences of incomplete preferences in multidimensional well-being comparisons. We have shown that incomplete preferences introduce an unavoidable normative choice into well-being measurement.

We provide an axiomatic characterization showing that any well-being measure satisfying four natural axioms must aggregate equivalent income bounds using a normative parameter α that reflects attitudes toward incompleteness. This parameter is unavoidable: no aggregation rule can produce a complete ranking without taking a stand on how unresolved trade-offs should be resolved. Equivalent income bounds reduce a multidimensional evaluation problem to a two-dimensional representation while making explicit the normative assumptions required to complete it.

Using survey data from 2,050 Dutch adults, we find that 26 percent of respondents explicitly indicate that they cannot rank their own life situation against plausible alternatives. This incompleteness emerges early in the elicitation process, persists across rounds, and is associated with longer response times, indicating that it reflects primarily genuine evaluative difficulty rather than measurement error.

The choice of α has substantial implications. Only 42 percent of individuals in the bottom decile under the lower-bound approach remain there under the upper-bound approach. Evaluating individuals at the lower bound of their interval systematically prioritizes respondents with the most incomplete preferences rather than those whose low well-being is precisely identified. At $\alpha = 0$, the bottom decile includes many individuals with relatively high incomes whose classification as worst-off arises primarily from wide intervals rather than low observed outcomes. Increasing the weight on upper bounds shifts the composition toward those with clearly identified low well-being. Lower-bound and upper-bound approaches also affect measured inequality: Gini coefficients differ by up to five points depending on the aggregation rule. There is no neutral default; any ranking or inequality measure inherently reflects a normative stance on how incompleteness is handled.

Different interpretations of the source of incompleteness may suggest different normative intuitions about how interval bounds should be aggregated. For example, incompleteness arising from fundamental incommensurability across life dimensions may support cautious aggregation, while incompleteness due to preference uncertainty or elicitation imprecision may call for alternative treatments. Our analysis deliberately does not take a stand on these interpretations. Instead, it shows that even when the source of incompleteness is left unspecified—or when multiple sources coexist—any attempt to produce complete welfare rankings must rely on an explicit normative choice. Disagreement about the origins of incompleteness therefore reinforces, rather than resolves, the need for transparent aggregation rules and sensitivity analysis.

These findings have implications for preference elicitation methods as well as for the design of prioritarian policies. While relatively few elicitation rounds suffice to rank most pairs of individuals, additional rounds sharpen the distinction between intrinsic incompleteness and measurement imprecision. Improving elicitation technology alone cannot substitute for explicit normative choices. On the policy side, one response is to target interventions toward individuals who satisfy the dominance criterion—worst-off under both lower-bound and upper-bound approaches—thereby avoiding dependence on a particular α, though at the cost of leaving some comparisons unresolved. A middle ground is to require dominance over a subinterval of $[0, 1]$ rather than the full range—for instance, $\alpha \in [0.25, 0.5]$—which narrows the normative commitment while still ranking more pairs than the full dominance criterion. A limitation of our analysis is that we treat incompleteness as exogenous; if individuals anticipate that wide intervals confer priority under low-α rules, strategic incentives to express incompleteness may arise.

More broadly, interval-valued welfare parameters arise in a variety of contexts. In classical revealed-preference settings, consumer surplus may be only partially identified (Willig, 1976), with bounds derived from limited knowledge of demand elasticities. In models of stochastic choice based on random utility maximization (McFadden, 1974), welfare measures may likewise be set-identified rather than point-identified, even when underlying preferences are assumed to be complete (see, e.g., Capéau et al., 2025). In program evaluation, treatment effects are frequently bounded in partially identified models (Manski, 1990, 2003), while in household economics, intra-household resource shares are often interval-valued in collective models (Chiappori and

Meghir, 2015; Cherchye et al., 2012; Browning et al., 2013). Converting such interval-valued welfare objects into point estimates or complete rankings requires normative choices analogous to the incompleteness attitude parameter α studied here.[31]

The central message is that incomplete preferences require explicit normative choices in well-being comparisons. Appeals to caution cannot avoid this: what appears cautious may instead privilege those with wide intervals over those whose low well-being is precisely identified. When preferences are incomplete, both the ranking of the worst-off and the measurement of overall inequality depend on how unresolved evaluative ambiguity is aggregated, highlighting its inescapable normative content.

[31] While we treat α as a normative parameter reflecting attitudes toward incompleteness, recent work explores alternative ways of weighting interval bounds under different assumptions about uncertainty regarding the location of welfare within the interval (Burone, 2026).

References

ADLER, M. D. AND O. F. NORHEIM (2022): *Prioritarianism in Practice*, Cambridge University Press.

ARROW, K. J. (1951): *Social Choice and Individual Values*, New York: Wiley.

ARROW, K. J. AND L. HURWICZ (1977): "An Optimality Criterion for Decision-making under Ignorance," in *Studies in Resource Allocation Processes*, Cambridge University Press, 463–471.

ATTEMA, A. E. AND W. B. BROUWER (2013): "In search of a preferred preference elicitation method: A test of the internal consistency of choice and matching tasks," *Journal of Economic Psychology*, 39, 126–140.

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.

BEWLEY, T. F. (2002): "Knightian Decision Theory. Part I," *Decisions in Economics and Finance*, 25, 79–110.

BROWNING, M., P.-A. CHIAPPORI, AND A. LEWBEL (2013): "Estimating Consumption Economies of Scale, Adult Equivalence Scales, and Household Bargaining Power," *The Review of Economic Studies*, 80, 1267–1303.

BUCHER-KOENEN, T., R. ALESSIE, A. LUSARDI, AND M. VAN ROOIJ (2025): "Fearless woman: Financial literacy, confidence, and stock market participation," *Management Science*, 71, 7414–7430.

BURONE, S. (2026): "Preference Refinements for Well-Being Comparisons under Incomplete Preferences," Tech. rep., (forthcoming).

BURONE, S. AND L. LEITNER (2025): "Correcting for Starting Point Bias in the Elicitation of Willingness to Pay for Health," Tech. rep., Herman Deleeck Centre for Social Policy, University of Antwerp.

BUTLER, D. J. AND G. C. LOOMES (2007): "Imprecision as an Account of the Preference Reversal Phenomenon," *The American Economic Review*, 97, 277–297.

CABEZA MARTÍNEZ, B. AND K. DECANCQ (2025): "How information about effort and luck shapes altruism of social preferences: a survey experiment: B. Cabeza Martínez and K. Decancq," *The Journal of Economic Inequality*, 23, 695–716.

Capéau, B., L. De Sadeleer, and S. Maes (2025): "Identifying the distribution of welfare from discrete choice," *Social Choice and Welfare*, 1–41.

Cherchye, L., B. De Rock, and F. Vermeulen (2012): "Married with Children: A Collective Labor Supply Model with Detailed Time Use and Intrahousehold Expenditure Information," *The American Economic Review*, 102, 3377–3405.

Chiappori, P.-A. and C. Meghir (2015): "Intrahousehold Inequality," in *Handbook of Income Distribution*, Elsevier, vol. 2, 1369–1418.

Da Costa, S., K. Decancq, M. Fleurbaey, and E. Schokkaert (Forthcoming): "Preference elicitation methods and equivalent income: an overview," *Reviews of Economic Literature*.

Deaton, A. and J. Muellbauer (1980): *Economics and Consumer Behavior*, Cambridge University Press.

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," *Handbook of Income Distribution*, 2, 67 – 140.

Decancq, K. and A. Nys (2021): "Non-parametric well-being comparisons," *European Economic Review*, 133, 103666.

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*, vol. 48, Cambridge University Press.

Fleurbaey, M. and E. Schokkaert (2013): "Behavioral welfare economics and redistribution," *American Economic Journal: Microeconomics*, 5, 180–205.

Gilboa, I. and D. Schmeidler (1989): "Maxmin Expected Utility with Non-unique Prior," *Journal of Mathematical Economics*, 18, 141–153.

Harrison, G., D. Ross, and J. T. Swarthout (Forthcoming): "Gender, Confidence, and the Mismeasure of Intelligence, Competitiveness and Literacy," *Journal of Political Economy*.

Hurwicz, L. (1951): "Optimality Criteria for Decision Making under Ignorance," Tech. rep., Cowles Commission discussion paper.

Karni, E. and M.-L. Vierø (2023): "Comparative incompleteness: measurement, behavioral manifestations and elicitation," *Journal of Economic Behavior & Organization*, 205, 423–442.

Krosnick, J. A. (1991): "Response strategies for coping with the cognitive demands of attitude measures in surveys," *Applied cognitive psychology*, 5, 213–236.

Lichtenstein, S. and P. Slovic (2006): *The construction of preference*, Cambridge University Press.

Manski, C. F. (1990): "Nonparametric Bounds on Treatment Effects," *The American Economic Review*, 80, 319–323.

——— (2003): *Partial Identification of Probability Distributions*, Springer.

McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, New York: Academic Press, 105–142.

Mirrlees, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 38, 175–208.

Nielsen, K. and L. Rigotti (2023): "Revealed incomplete preferences," *Available at SSRN 4622145*.

Parfit, D. (1984): *Reasons and Persons*, Oxford, UK: Oxford University Press.

Stantcheva, S. (2023): "How to run surveys: A guide to creating your own identifying variation and revealing the invisible," *Annual Review of Economics*, 15, 205–234.

Sturgis, P. and I. Brunton-Smith (2023): "Personality and survey satisficing," *Public Opinion Quarterly*, 87, 689–718.

Weymark, J. A. (1981): "Generalized Gini inequality indices," *Mathematical social sciences*, 1, 409–430.

Willig, R. D. (1976): "Consumer's surplus without apology," *The American Economic Review*, 66, 589–597.

APPENDIX

A Proof of Theorem 1

Theorem 1. *A well-being measure* $w : \mathcal{L} \times \mathcal{P} \to \mathbb{R}_+$ *satisfies Bound Invariance, Monotonicity in Bounds,* α*-balanced Invariance (for some* $\alpha \in [0, 1]$*), and Normalization if and only if, for all* $\ell \in \mathcal{L}$ *and* $P \in \mathcal{P}$,

$$w(\ell, P) = (1 - \alpha)\, \underline{y}^\star(\ell, P) + \alpha\, \overline{y}^\star(\ell, P).$$

Proof. We prove necessity; sufficiency follows by direct verification.

Step 1. Reduction to the bounds. Let

$$u := \underline{y}^\star(\ell, P), \qquad v := \overline{y}^\star(\ell, P).$$

By Bound Invariance (Axiom 1), well-being depends on (ℓ, P) only through (u, v). Hence there exists a function

$$W : \mathbb{R}^2 \to \mathbb{R}$$

such that $w(\ell, P) = W(u, v)$. Continuity of the bounds implies that W is continuous, and Monotonicity in Bounds (Axiom 2) implies that W is weakly increasing in each argument.

Step 2. Balanced invariance. Axiom 3 implies that for any (u, v) and (u', v'),

$$\alpha(v - v') = (1 - \alpha)(u' - u) \quad \Longrightarrow \quad W(u, v) = W(u', v').$$

Rearranging yields

$$(1 - \alpha)u + \alpha v = (1 - \alpha)u' + \alpha v'.$$

Thus, well-being is constant on each set

$$\{(u, v) \in \mathbb{R}^2 : (1 - \alpha)u + \alpha v = s\}, \quad s \in \mathbb{R}.$$

Step 3. Functional representation. Define $s := (1 - \alpha)u + \alpha v$. Since $W(u, v)$ is constant for fixed s, there exists a function $F : \mathbb{R} \to \mathbb{R}$ such that

$$W(u, v) = F\big((1 - \alpha)u + \alpha v\big).$$

Continuity of W implies continuity of F, and monotonicity of W implies that F is strictly increasing. This representation follows from standard results on continuous functions invariant under translations along a fixed direction (see Aczél 1966, p. 58).

Step 4. Normalization. By Normalization (Axiom 4), $W(y, y) = y$ for all y, which implies

$$F(y) = y \quad \text{for all } y.$$

Hence

$$W(u, v) = (1 - \alpha)u + \alpha v,$$

and therefore

$$w(\ell, P) = (1 - \alpha)\, \underline{y}^{\star}(\ell, P) + \alpha\, \overline{y}^{\star}(\ell, P).$$

□

B Adaptive Bisectional Dichotomous Choice Method

This Appendix provides additional details on the Adaptive Bisectional Dichotomous Choice (ABDC) algorithm. Appendix B.1 presents the pseudocode, and Appendix B.2 provides a screenshot of the dichotomous choice interface.

B.1 Pseudocode

Algorithm 1 Adaptive Bisectional Dichotomous Choice (ABDC)

Require: Respondent's life situation $\ell = (y, z)$, reference bundle $\overline{z}$, maximum elicitation rounds R, precision ϵ

Initialize $\underline{y}^\star_0 \leftarrow 0$, $\overline{y}^\star_0 \leftarrow y$
for $r = 1$ to R **do**
    Set $y^\star_r \leftarrow (\underline{y}^\star_{r-1} + \overline{y}^\star_{r-1})/2$
    Present choice between ℓ and $\ell' = (y^\star_r, \overline{z})$
    if $\ell' \in UC(\ell, P)$ **then**
        $\overline{y}^\star_r \leftarrow y^\star_r$, $\underline{y}^\star_r \leftarrow \underline{y}^\star_{r-1}$
    else if $\ell' \in LC(\ell, P)$ **then**
        $\underline{y}^\star_r \leftarrow y^\star_r$, $\overline{y}^\star_r \leftarrow \overline{y}^\star_{r-1}$
    else if Respondent selects "Don't know" **then**
        break
    end if
    if $\overline{y}^\star_r - \underline{y}^\star_r \leq \epsilon$ **then**
        break
    end if
end for
return $[\underline{y}^\star_r, \overline{y}^\star_r]$

B.2 Screen Shot

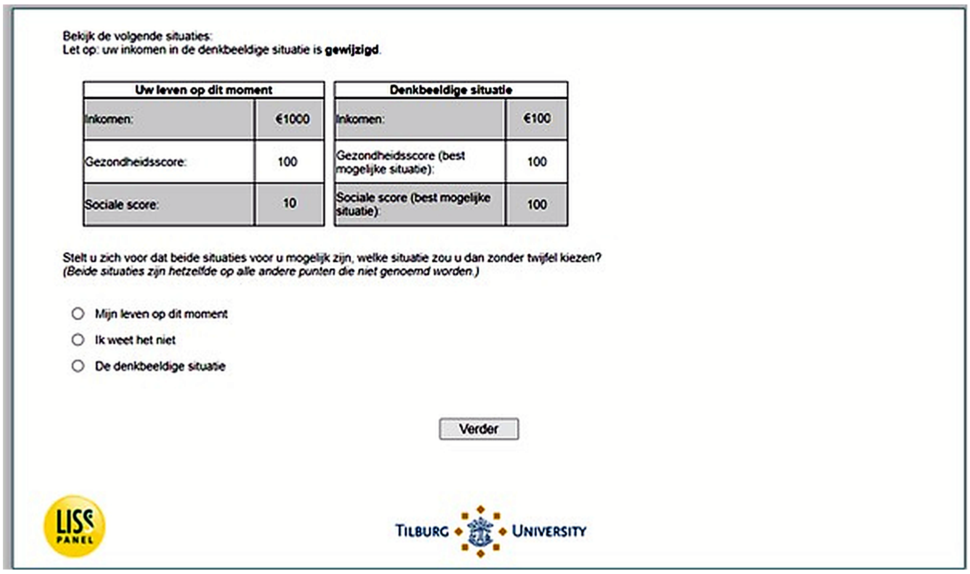

Figure 7 Screenshot of a Dichotomous Choice of the ABDC

Translation of Figure 7:

Look at the following situations:
Please note: your income in the imaginary situation has **changed**.

Your life at this moment		**Imaginary situation**	
Income:	€1000	Income:	€100
Health score:	100	Health score: (best possible situation)	100
Social score:	10	Social score: (best possible situation)	100

Imagine that both situations are possible for you, which situation would you choose without a doubt? (Both situations are the same on all other points not mentioned.)

- ○ **My life at this moment**
- ○ **I don't know**
- ○ **The imaginary situation**

C Additional Figures and Tables

C.1 Distribution of Well-Being Dimensions and Adjustments

Figures 8, 9, and 10 display the distributions of income, health, and social relations, respectively. Each figure shows both the pre-loaded distribution and the distribution after adjustment.[32] The income distribution is right-skewed, with most mass concentrated at lower and middle values and a long upper tail. The distributions of health and social relations are less skewed, with greater mass at higher values and thinner lower tails.

Only a small share of respondents adjusted the pre-loaded values. Adjustments are generally modest and primarily shift probability mass toward the extremes. Table 4 reports summary statistics. Income was adjusted most frequently: 44 percent of respondents revised their pre-loaded values, compared to 21 percent for health and 13 percent for social relations. Across all three dimensions, the mean signed adjustment is positive, indicating that respondents tend to revise values upward. Among those who made adjustments, revisions are small, with median absolute changes of 63 euros for income, 8 points for health and 12 points for social relations. The distributions of adjustments are right-skewed, explaining why means exceed medians.

[32] Appendix D.2 describes the construction of the pre-loaded variables.

Table 4 Descriptive Statistics of Adjustments in Well-Being Dimensions

Dimension	Adjustment	Mean	Median	Min	Max	SD	*N*
Income	Signed	43.69	0.50	-4,362	23,000	1,337.8	909
	Absolute	370.1	63.0	0.2	23,000	1,286.3	909
Health	Signed	4.07	8	-76	34	15.9	423
	Absolute	13.6	12	1	76	9.2	423
Social	Signed	4.92	8	-72	56	17.2	264
	Absolute	14.8	12	1	72	10.1	264

Notes: Each row reports statistics for respondents who adjusted the pre-loaded value in that dimension. Signed adjustments in the first row; absolute adjustments in the second.

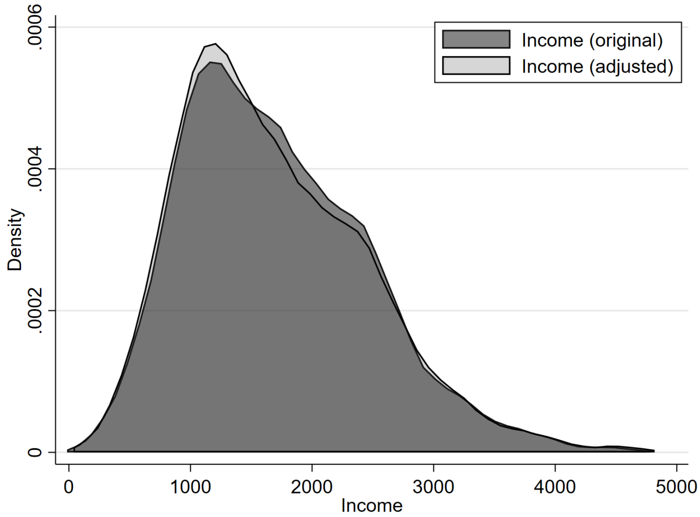

Figure 8 Original (Pre-loaded) and Adjusted Distribution of Income
Notes: Weighted baseline sample (*N* = 2,050). See Appendices E.1–E.2 for robustness.

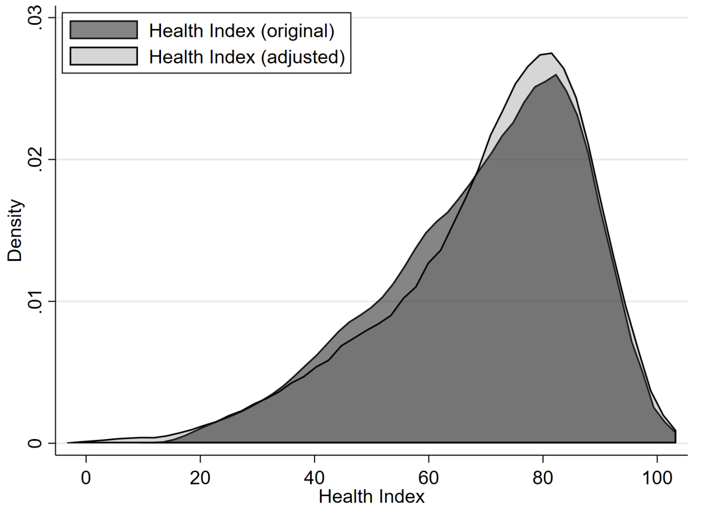

Figure 9 Original (Pre-loaded) and Adjusted Distribution of the Health Index
Notes: Weighted baseline sample (N = 2,050). See Appendices E.1–E.2 for robustness.

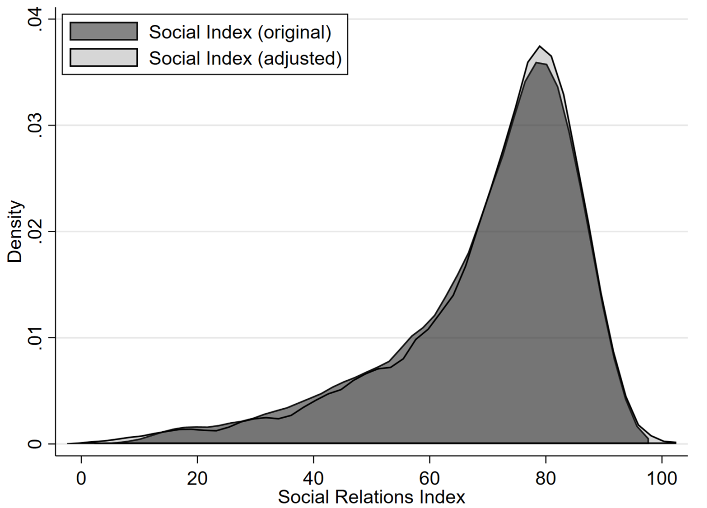

Figure 10 Original (Pre-loaded) and Adjusted Distribution of the Social Relations Index
Notes: Weighted baseline sample (N = 2,050). See Appendices E.1–E.2 for robustness.

C.2 Cumulative Prevalence of “Don’t know” Responses by Gender

Figure 3 illustrates how the cumulative prevalence of “Don’t know” responses evolves across the ten rounds of the ABDC elicitation, disaggregated by gender. Two patterns stand out. First, women end the procedure with a lower overall cumulative prevalence than men, consistent with the regression results in Table 2. Second, women select “Don’t know” slightly more often in the first two rounds, suggesting that female respondents express initial caution when confronted with unfamiliar trade-offs but resolve incomparabilities more quickly as the elicitation procedure progresses. Men, by contrast, select “Don’t know” responses more gradually, yet persistently, resulting in a higher cumulative prevalence.

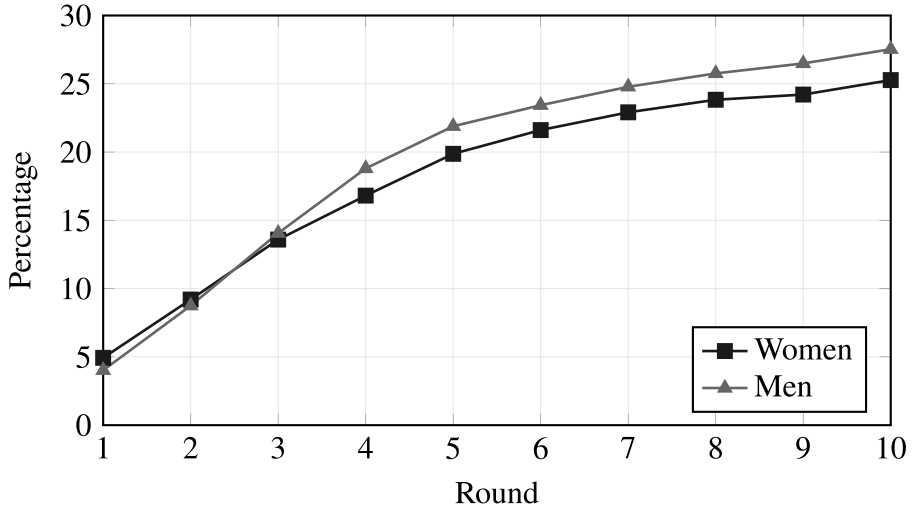

Figure 11 Cumulative prevalence of “Don’t know” responses by elicitation round and by gender
Notes: Unweighted baseline sample (N = 2,050).

C.3 Response Time

Figure 12 reports the average time respondents spent viewing the screen in each round of the ABDC procedure, disaggregated by response category. Two main patterns emerge. First, the initial round(s) are typically associated with longer viewing times, which can be interpreted as evidence that respondents require more time at the beginning to understand the task and unfold their preferences. Once the procedure becomes familiar, subsequent rounds tend to be completed more quickly. Second, across all rounds, the average time spent on the screen is systematically higher when respondents select “Don’t know”. This pattern is consistent with the interpretation that such responses reflect genuine incompleteness of preferences rather than lack of engagement with the questionnaire.

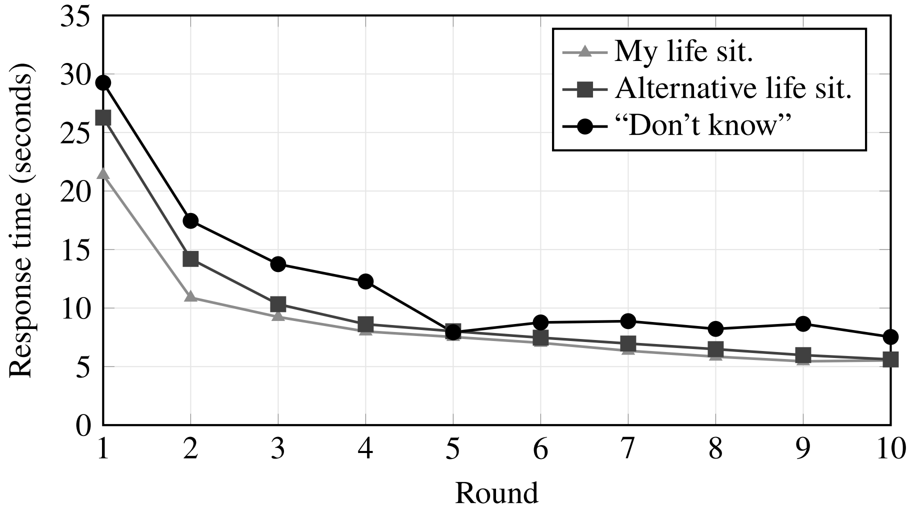

Figure 12 Mean response time by response category and elicitation round.
Notes: Weighted baseline sample (N = 2,050).

C.4 Distribution of Incompleteness Measure $\Delta y^\star$

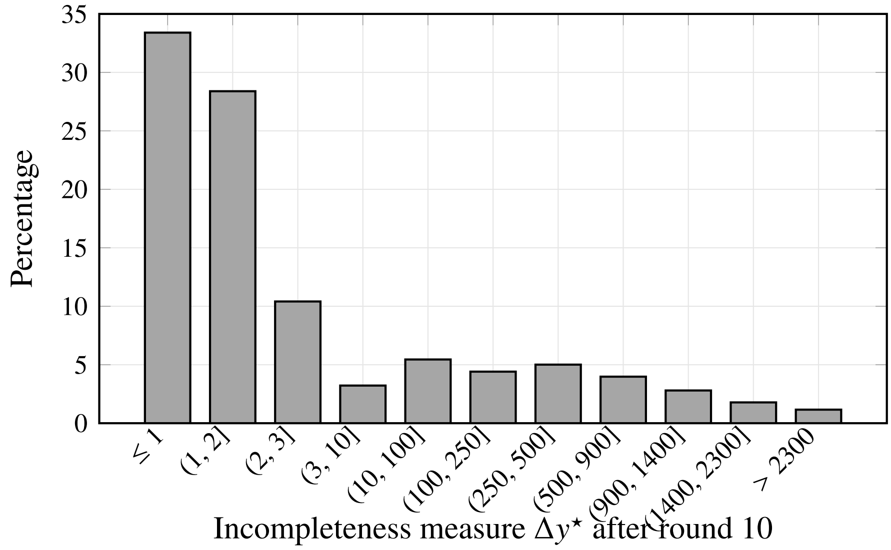

Figure 13 Distribution of Incompleteness Measure after Ten Elicitation Rounds
Notes: Weighted baseline sample (N = 2,050).

C.5 Distribution of Pairwise Combinations

There are 2,050 respondents in the sample, yielding 2,100,225 unique pairs of respondents that can be compared using the well-being criteria, discussed in Section 2.2. Table 5 shows the breakdown of these pairs by the number of respondents in each pair who used the "Don't know" response to terminate the algorithm, indicating intrinsic incompleteness. Approximately 46 percent of pairs include at least one respondent who terminated with "Don't know": 7 percent involve two such respondents and 39 percent involve exactly one."

Table 5 Distribution of Pairwise Combinations by Number of "Don't Know" Responses

# "Don't Know" Responses	# Pairs	Percentage
0 per pair	1,134,740	54.03%
1 per pair	818,330	38.96%
2 per pair	147,155	7.01%
Total	2,102,225	100.0%

Notes: Weighted baseline sample (N = 2,050).

C.6 Critical Values for Ranking Reversals According to Hurwicz Criterion

Figure 14 focuses on pairs of respondents who are not ranked according to the dominance criterion after elicitation round 10 (11.0 percent of all pairs), meaning that their relative rankings based on the lower bound ($\alpha = 0$) and upper bound ($\alpha = 1$) differ. The histogram shows the distribution of $\alpha^\star$, the critical value of α at which the two individuals' Hurwicz equivalent incomes w_α are equal. At values of α above or below $\alpha^\star$, the well-being ranking between the pair reverses. The distribution of $\alpha^\star$ spans the unit interval, with greater mass concentrated toward lower values. This pattern indicates that most ranking reversals occur at low levels of α, because differences between individuals' lower bounds are generally smaller in magnitude than the differences between their upper bounds.

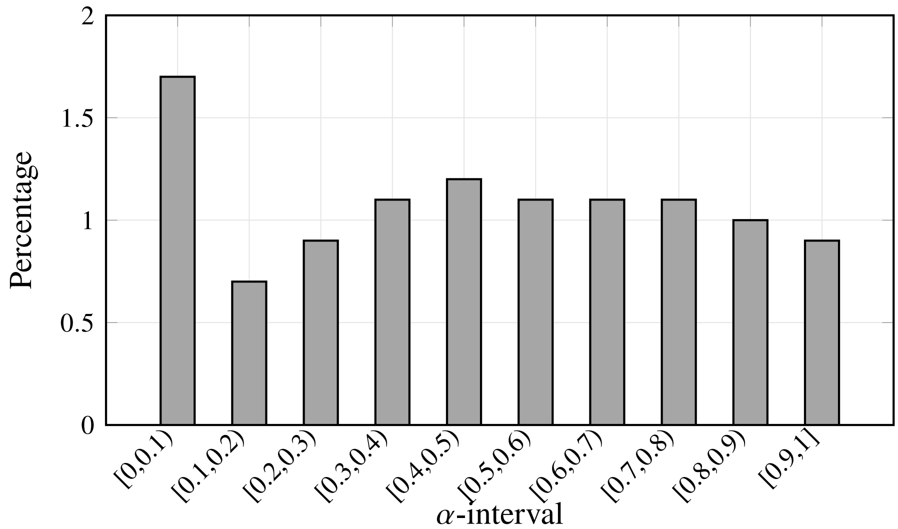

Figure 14 Distribution of Critical Values $\alpha^\star$ for Ranking Reversals
Notes: The figure shows the distribution of $\alpha^\star$ values for pairs of respondents whose rankings differ between the lower bound ($\alpha = 0$) and upper bound ($\alpha = 1$) after elicitation round 10. For each pair, $\alpha^\star$ is the critical value at which their Hurwicz equivalent incomes are equal and their ranking according to the Hurwicz criterion reverses. Weighted baseline sample ($N = 2{,}050$).

C.7 Transition Matrix: Lower and Upper Bounds

Table 6 presents a transition matrix by deciles of the Hurwicz equivalent income measure, $w_\alpha(\ell, P)$, comparing the lower bound ($\alpha = 0$) with the upper bound ($\alpha = 1$). The concentration of mass around the diagonal indicates that for most respondents, the two rankings differ by no more than one decile, reflecting the fact that most individuals have relatively narrow equivalent income intervals. As a summary measure of concordance, the Spearman rank correlation between the lower and upper bound rankings is 0.85.

However, there are notable exceptions. Among individuals classified in the worst-off decile under $\alpha = 0$, approximately 11 percent are classified in the top decile under $\alpha = 1$. More broadly, the overlap between the two rankings is only 42 percent in the bottom decile and 75 percent in the top decile, highlighting substantial sensitivity of the rankings to the choice of α, particularly at the lower end of the well-being distribution.

Table 6 Transition Matrix of Hurwicz Equivalent Income: $\alpha = 0$ versus $\alpha = 1$

	$\alpha = 1$										
1	42.4	5.9	6.3	5.9	9.8	7.3	2.4	3.9	5.4	10.7	205
2	57.6	29.8	4.4	5.4	1.5	0.5	1.0	0.0	0.0	0.0	205
3	0.0	64.4	22.0	2.4	4.9	3.9	2.0	0.0	0.0	0.5	205
4	0.0	0.0	67.3	20.0	5.9	2.4	1.5	2.9	0.0	0.0	205
5	0.0	0.0	0.0	66.3	23.9	5.9	2.9	0.5	0.5	0.0	205
6	0.0	0.0	0.0	0.0	54.2	38.1	2.4	1.0	0.5	3.9	205
7	0.0	0.0	0.0	0.0	0.0	42.0	51.2	2.4	2.0	2.4	205
8	0.0	0.0	0.0	0.0	0.0	0.0	36.6	53.2	7.3	2.9	205
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	36.1	59.5	4.4	205
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	24.9	75.1	205
N	205	205	205	205	205	205	205	205	205	205	2,050

Notes: Each cell reports the percentage of respondents transitioning between deciles under $\alpha = 0$ (rows) and $\alpha = 1$ (columns). Both rows and columns sum to 100 percent. Ties were broken at random to ensure equal numbers of respondents in each decile. Weighted baseline sample ($N = 2{,}050$).

C.8 Alternative Specifications

Table 7 reports a set of robustness checks for the results on determinants of preferences incompleteness. Models (1) and (3) correspond to the Probit and OLS specifications presented in the main text (Table 2). Columns labelled (') use a slightly different specification: age is included as a continuous variable, only unemployment is added as a control for occupation, and education is excluded. The main results remain unchanged.

Column (2) estimates the same outcome as in column (1) using a logit model instead of a probit model. The results are very similar. Column (4) uses an alternative outcome measure that captures the degree of incompleteness, taking into account at which stage respondents answered "Don't know". The ordered probit estimates confirm the main findings.

Across all specifications, greater distance to the reference situation, longer response time, and finding the questions unclear are consistently associated with higher levels of incompleteness. This supports the interpretation that "Don't know" answers mainly reflect genuine uncertainty in preferences rather than random noise.

Table 7 Determinants of Incompleteness: Robustness Across Specifications

Model	(1)	(1')	(2)	(2')	(3)	(3')	(4)	(4')
Dep. Variable	"Don't know"				$\Delta y^{\star}$		"Incompleteness degree"	
Logarithm income	0.107	0.108	0.181	0.194	0.0112	0.0240	-0.274***	-0.290***
	(0.0740)	(0.0708)	(0.133)	(0.126)	(0.00776)	(0.0164)	(0.0665)	(0.0625)
Health index	-0.00572**	-0.00626**	-0.0102**	-0.0113**	-0.000528*	-0.00142**	-0.00598***	-0.00647***
	(0.00244)	(0.00246)	(0.00440)	(0.00447)	(0.000283)	(0.000578)	(0.00209)	(0.00209)
Social relations index	-0.00130	-0.00184	-0.00192	-0.00270	-0.000345	-0.000462	-0.00169	-0.00235
	(0.00256)	(0.00253)	(0.00454)	(0.00448)	(0.000278)	(0.000591)	(0.00205)	(0.00206)
Female	-0.192**	-0.192**	-0.321**	-0.335**	-0.00962	-0.0413**	-0.141**	-0.143**
	(0.0791)	(0.0791)	(0.143)	(0.143)	(0.00815)	(0.0174)	(0.0687)	(0.0686)
Above Median Age	-0.333***		-0.606***		-0.0486***		-0.353***	
	(0.104)		(0.189)		(0.0104)		(0.0877)	
Age		-0.0181***		-0.0329***		-0.00439***		-0.0184***
		(0.00303)		(0.00548)		(0.000701)		(0.00267)
University degree	0.208*		0.377*		-0.00194		0.122	
	(0.109)		(0.194)		(0.0109)		(0.0948)	
Pensioner	-0.247**		-0.397*		-0.0146		-0.242**	
	(0.116)		(0.214)		(0.0125)		(0.100)	
Occupation: paid employee = 1	-0.0556		-0.0495		0.00544		-0.0348	
	(0.104)		(0.189)		(0.0100)		(0.0856)	
Occupation: unemployed = 1		0.0897		0.174		0.0171		-0.115
		(0.239)		(0.437)		(0.0577)		(0.187)
Openness	-0.0236	0.00121	-0.0307	0.0136	0.0105	0.00247	0.0434	0.0507
	(0.0824)	(0.0812)	(0.148)	(0.145)	(0.00852)	(0.0183)	(0.0714)	(0.0707)
Neurotic	0.0296	0.00773	0.0630	0.0219	0.00869	-0.000803	0.00947	-0.00932
	(0.0634)	(0.0635)	(0.115)	(0.116)	(0.00633)	(0.0146)	(0.0532)	(0.0538)
Conscientiousness	-0.178*	-0.169*	-0.315*	-0.286	-0.00723	-0.0380*	-0.119	-0.106
	(0.0996)	(0.0989)	(0.180)	(0.178)	(0.0102)	(0.0227)	(0.0850)	(0.0850)
Agreeableness	-0.0970	-0.0975	-0.188	-0.192	-0.0102	-0.0233	-0.0983	-0.0969
	(0.0811)	(0.0810)	(0.148)	(0.146)	(0.00809)	(0.0183)	(0.0674)	(0.0676)
Extraversion	-0.0408	-0.0392	-0.0829	-0.0810	-0.00146	-0.0106	-0.0384	-0.0391
	(0.0599)	(0.0591)	(0.110)	(0.107)	(0.00572)	(0.0137)	(0.0513)	(0.0508)
Did the questions make you think?	0.323***	0.310***	0.565***	0.536**	0.0112	0.0581***	0.211**	0.196**
	(0.118)	(0.118)	(0.219)	(0.220)	(0.0107)	(0.0224)	(0.0941)	(0.0943)
Did you find the questions unclear?	0.487***	0.513***	0.857***	0.911***	0.0525***	0.115***	0.396***	0.424***
	(0.139)	(0.139)	(0.242)	(0.244)	(0.0198)	(0.0353)	(0.120)	(0.120)
Did you find the questions interesting?	-0.0410	-0.0551	-0.0386	-0.0574	-0.0153	-0.000878	-0.0464	-0.0650
	(0.144)	(0.145)	(0.268)	(0.270)	(0.0153)	(0.0294)	(0.118)	(0.119)
Time last round	0.165***	0.166***	0.293***	0.296***	0.0283***	0.0463***	0.158***	0.160***
	(0.00950)	(0.00957)	(0.0190)	(0.0194)	(0.00140)	(0.00146)	(0.00659)	(0.00671)
Threshold 1							-1.542***	-2.509***
							(0.539)	(0.551)
Threshold 2							-1.340**	-2.307***
							(0.533)	(0.544)
Threshold 3							-1.275**	-2.242***
							(0.532)	(0.544)
Threshold 4							-1.207**	-2.174***
							(0.532)	(0.544)
Threshold 5							-1.125**	-2.092***
							(0.532)	(0.544)
Threshold 6							-0.952*	-1.919***
							(0.533)	(0.545)
Threshold 7							-0.674	-1.640***
							(0.533)	(0.545)
Threshold 8							-0.242	-1.206**
							(0.534)	(0.547)
Threshold 9							0.412	-0.550
							(0.538)	(0.548)
Constant	-1.285**	-0.486	-2.196**	-0.810	-0.0885	0.287**		
	(0.618)	(0.631)	(1.098)	(1.118)	(0.0657)	(0.146)		
Estimation method	Probit	Probit	Logit	Logit	OLS	OLS	Oprobit	Oprobit
N	1,953	1,953	1,953	1,953	1,953	1,953	1,953	1,953
(Pseudo) R^2	0.309	0.309	0.308	0.309	0.474	0.352	0.191	0.191

Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Notes: In models (4) and (4'), the dependent variable "Incompleteness degree" ranges from 0 to 9 and equals the maximum possible rounds minus the rounds answered. A value of 0 means all rounds were answered; a value of 9 means the respondent answered 'I don't know' in the first round.

C.9 Distribution of Hurwicz equivalent incomes

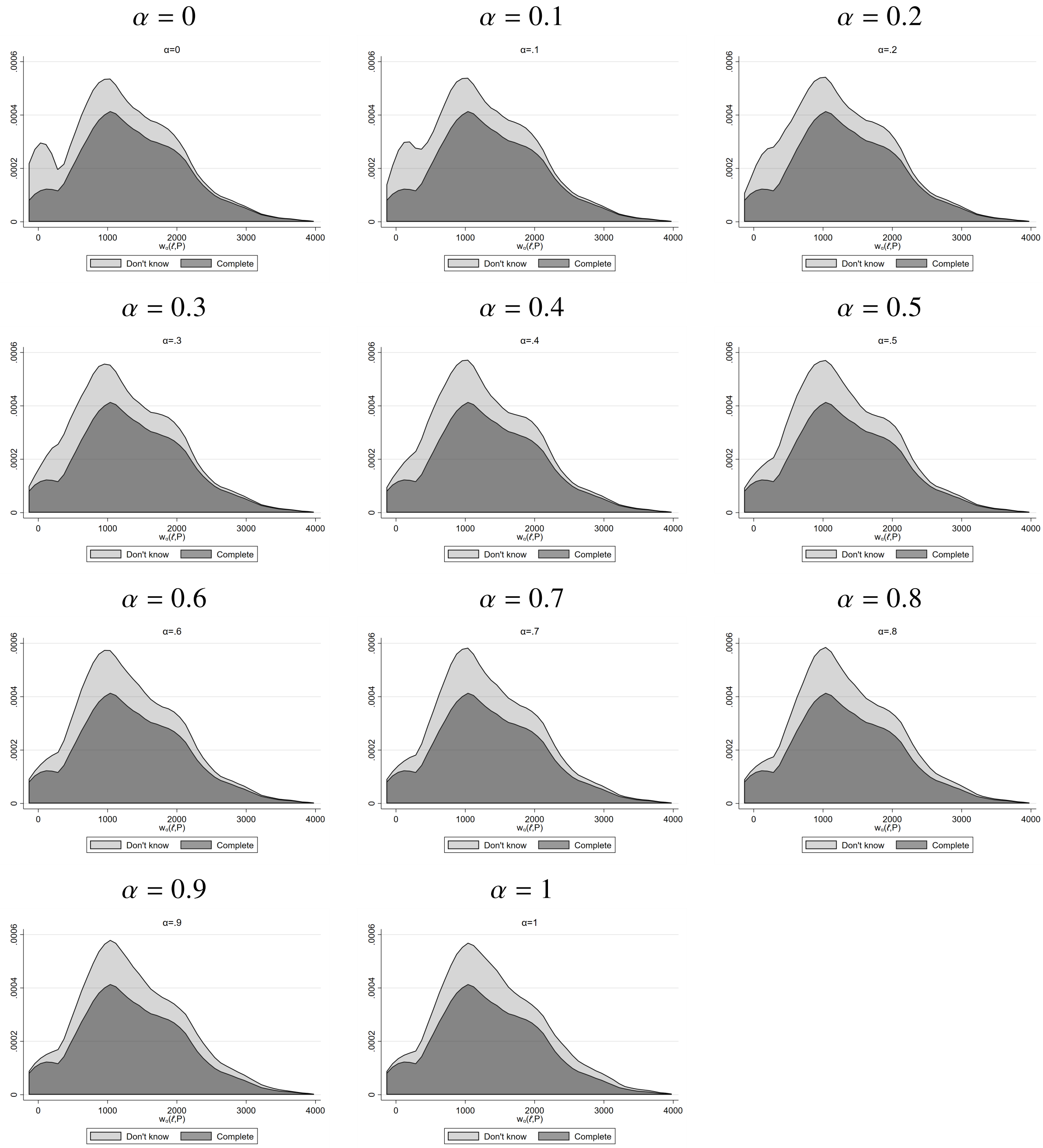

Figure 15 Hurwicz equivalent incomes across α values
Notes: Kernel density of Hurwicz equivalent incomes by α. Epanechnikov kernel, bandwidth 138.5. Weighted baseline sample (N = 2,050). See Appendices E.1–E.2 for robustness.

C.10 S-Gini Inequality Indices Across α Values

Let $w_\alpha(\ell_i, P_i)$ denote the Hurwicz equivalent income of individual $i = 1, \ldots, N$ as defined by Equation (3). Individuals are indexed so that $w_\alpha(\ell_1, P_1) \leq w_\alpha(\ell_2, P_2) \leq \cdots \leq w_\alpha(\ell_N, P_N)$, with $i = 1$ denoting the worst-off and $i = N$ the best-off individual in terms of Hurwicz equivalent income. The S-Gini family of inequality indices is defined as:

$$I_\delta(W_\alpha) = 1 - \sum_{i=1}^{N} \left[\left(\frac{N-i+1}{N} \right)^{\delta} - \left(\frac{N-i}{N} \right)^{\delta} \right] \frac{w_\alpha(\ell_i, P_i)}{\mu_\alpha}, \tag{4}$$

where $\mu_\alpha = \frac{1}{N} \sum_{j=1}^{N} w_\alpha(\ell_j, P_j)$ is the mean of the well-being distribution $W_\alpha = (w_\alpha(\ell_i, P_i))_{i=1}^{N}$. The normative parameter $\delta > 1$ governs the bottom-sensitivity of the measure, determining the weight placed on the lower tail of the well-being distribution. This rank-weighted sum places progressively greater emphasis on lower-ranked individuals as δ increases. When $\delta = 2$, the standard Gini coefficient is obtained. As $\delta \to \infty$, the measure converges to a Rawlsian perspective where all weight is placed on the worst-off individual.[33]

Table 8 reports how different members of the S-Gini family, given by $\delta \in \{2, 5, 10\}$, depend on the incompleteness attitude $\alpha \in \{0, 0.25, 0.5, 0.75, 1\}$ across the ten rounds of the ABDC procedure. Several patterns emerge from these results. First, inequality is systematically higher when measured using the lower bound ($\alpha = 0$) compared to the upper bound ($\alpha = 1$), in line with the pattern shown in Figure 5. Second, the gap between inequality measured with the lower and upper bounds is substantially larger for higher bottom-sensitivity parameters (higher δ). For the standard Gini coefficient ($\delta = 2$), the difference between $\alpha = 0$ and $\alpha = 1$ is 5.2 Gini points, whereas for the most bottom-sensitive measure ($\delta = 10$), this gap reaches 9.2 Gini points by round 10. Third, inequality measures stabilize after the initial rounds of the ABDC procedure, with most variation occurring between rounds 1–5 and relative stability thereafter. This convergence pattern holds across all combinations of α and δ, suggesting that the distributional properties become well-established early in the ABDC elicitation procedure.

[33] Weymark (1981) provides an axiomatic characterization.

Table 8 Bounds on Well-Being Inequality: *S*-Gini Index

α	Round 1	2	3	4	5	6	7	8	9	10
	Bottom-Sensitivity Parameter: δ = 2 (Gini)									
0	42.8	40.5	40.0	39.8	39.8	39.8	39.8	39.8	39.8	39.8
0.25	36.7	37.0	37.2	37.3	37.3	37.3	37.3	37.3	37.4	37.4
0.5	33.8	34.9	35.4	35.5	35.6	35.7	35.7	35.7	35.7	35.7
0.75	32.4	33.7	34.3	34.6	34.7	34.8	34.8	34.8	34.8	34.8
1	31.6	33.0	33.8	34.2	34.4	34.5	34.5	34.6	34.6	34.6
	Bottom-Sensitivity Parameter: δ = 5									
0	77.3	73.3	72.3	71.9	71.8	71.8	71.7	71.7	71.7	71.7
0.25	65.2	65.9	66.3	66.4	66.5	66.5	66.6	66.6	66.6	66.6
0.5	59.0	61.7	62.7	63.2	63.4	63.5	63.6	63.6	63.6	63.6
0.75	56.0	59.2	60.9	61.6	62.0	62.1	62.2	62.3	62.3	62.3
1	54.3	57.7	59.8	60.8	61.4	61.6	61.8	61.8	61.9	61.9
	Bottom-Sensitivity Parameter: δ = 10									
0	82.8	86.1	85.9	85.8	85.7	85.7	85.7	85.7	85.7	85.7
0.25	78.4	80.2	81.0	81.3	81.5	81.6	81.7	81.7	81.7	81.7
0.5	71.1	75.2	77.1	77.9	78.3	78.5	78.6	78.6	78.7	78.6
0.75	67.5	72.2	74.9	76.2	76.9	77.2	77.3	77.4	77.5	77.4
1	65.4	70.3	73.5	75.3	76.2	76.6	76.9	77.0	77.0	77.0

Notes: *S*-Gini indices of Hurwicz equivalent incomes by α and δ. Columns indicate ABDC elicitation round. Weighted baseline sample (N = 2,050). See Appendices E.1–E.2 for robustness.

D Sample and Variable Construction

This appendix provides additional details on the data. Section D.1 describes sample selection and the derivation of sample weights. Section D.2 explains the construction of the health and social relations indices.

D.1 Sample Selection and Weights

Table 9 summarizes the sample selection procedure. From the active LISS panel, which is based on a random draw of Dutch-speaking households from the Dutch population register, 3,269 individuals aged 25 to 75 were invited to participate. Of these, 2,543 agreed, forming the raw sample (response rate: 77.8 percent).

We apply several exclusion criteria. First, a randomly selected subgroup of approximately 10 percent completed a different version of the questionnaire and were excluded. Second, we drop 21 incomplete responses. Third, 5 respondents who adjusted their income to zero were excluded. Fourth, 50 respondents reported inconsistent preferences: they preferred their current life ℓ to $\ell_0 = (y, \bar{z})$ in the training question, but preferred $\ell_1 = (y/2, \bar{z})$ to ℓ in the first ABDC round, violating transitivity or monotonicity in income.[34] Fifth, 29 respondents were excluded based on comments indicating misunderstanding or lack of engagement.[35] This yields a full sample of 2,186 respondents.

Next, following Stantcheva (2023), we exclude 136 respondents for speeding.[36] The resulting baseline sample contains 2,050 respondents.

[34] Since ℓ_0 and ℓ_1 differ only in income, monotonicity implies $\ell_0 \; P \; \ell_1$. Combined with $\ell_1 \; P \; \ell$ and $\ell \; P \; \ell_0$, transitivity would require $\ell_1 \; P \; \ell_0$—a contradiction.

[35] For example, respondent 806707 wrote: *"The contrast offered is flawed. If my income were halved, both scores could not equal 100,"* suggesting the hypothetical nature of the scenarios was misunderstood. Respondent 809481 stated: *"What good is an imaginary situation if the reality of my illness never makes it possible?"* indicating difficulty engaging with counterfactual scenarios. Other excluded respondents indicate similar issues or expressed frustration suggesting limited engagement.

[36] Respondents were classified as speeders if: (i) they completed r ABDC rounds in less than r seconds; (ii) they spent less than 5 seconds reading the ABDC instructions; or (iii) they spent less than 10 seconds reading the instructions and completing the training question comparing ℓ and $\ell_0 = (y, \bar{z})$.

Table 9 Sample Selection

	N
Invited respondents	3,269
Declined participation	−726
Raw sample	2,543
Experimental treatment	−252
Incomplete responses	−21
Income adjusted to zero	−5
Inconsistent preferences	−50
Flagged by comments	−29
Full sample	2,186
Speeding	−136
Baseline sample	2,050

The panel structure of LISS allows us to adjust for selection on observables into the baseline sample. We construct inverse-probability weights based on invited respondents' estimated likelihood of appearing in the baseline sample, obtained from the logit model in Table 11. Older invited respondents, pensioners, and men are more likely to appear in the baseline sample.

Table 10 shows the effect of applying these inverse-probability weights on the age and gender composition of different samples. Weighting brings the distribution closer to that of the invited population, correcting for the overrepresentation of older respondents and men.

Table 10 Age Composition of Samples by Gender

	Invited ($N = 3,269$)			Full Sample ($N = 2,186$)			Unweighted Baseline ($N = 2,050$)			Weighted Baseline ($N = 2,050$)		
Age	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female
25–35	13.7	5.4	8.3	10.8	4.4	6.4	9.8	4.0	5.8	13.2	5.0	8.2
35–45	15.2	6.6	8.6	12.0	5.6	6.4	11.1	5.3	5.8	13.8	6.0	7.8
45–55	18.4	8.1	10.3	17.2	7.9	9.3	17.1	7.8	9.3	18.6	8.3	10.3
55–65	23.2	11.0	12.2	25.4	11.9	13.5	25.7	12.0	13.7	23.5	10.8	12.7
65–75	29.5	14.7	14.8	34.6	17.6	17.0	36.3	18.5	17.7	30.9	15.3	15.6
Total	100	45.8	54.2	100	47.5	52.5	100	47.7	52.3	100	45.4	54.6

The weighted baseline sample of 2,050 respondents is used in the main text. The unweighted baseline sample is used in the sensitivity analysis in Appendix E.1 and the unweighted full sample of 2,186 respondents in Appendix E.2.

Table 11 Selection into Baseline Sample

	Coefficient	
Female	-0.330*	(0.200)
Age 35–44	0.300	(0.219)
Age 45–54	0.531**	(0.217)
Age 55–64	1.315***	(0.228)
Age 65+	1.832***	(0.332)
Female × Age 35–44	-0.216	(0.277)
Female × Age 45–54	0.103	(0.273)
Female × Age 55–64	0.145	(0.290)
Female × Age 65+	-0.405	(0.328)
Married	0.141	(0.095)
Occupation (ref: employed)		
Family business	-0.059	(0.525)
Self-employed	0.070	(0.202)
Job seeker	0.624	(0.441)
First-time job seeker	0.318	(1.195)
Exempted from job seeking	-0.433	(0.675)
Student	0.018	(0.372)
Homemaker	0.294	(0.182)
Pensioner	0.705***	(0.256)
Work disability	0.000	(0.213)
Unpaid work	-1.093	(0.856)
Voluntary work	0.099	(0.426)
Other occupation	-0.162	(0.616)
Education (ref: primary)		
Intermediate secondary	0.206	(0.345)
Higher secondary	0.398	(0.367)
Intermediate vocational	0.146	(0.333)
Higher vocational	0.222	(0.332)
University	0.415	(0.345)
Other education	-0.469	(0.448)
Constant	0.209	(0.366)
Estimation method	Logit	
N	3,269	
Pseudo R^2	0.11	

Note: Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

D.2 Index Construction

D.2.1 Health Index

The Health index was constructed based on data from the Health Core Module of the LISS Panel. An index capturing both physical and mental health, composed of 10 sub-dimensions (listed in Figure 12) was pre-computed.[37]

The index ranges from 0 (worst) to 100 (best). Construction proceeded in three steps: (i) rescaling all sub-dimensions to 0–10; (ii) estimating weights via a regression of self-perceived health (1–5 scale) on the 10 sub-dimensions and demographic controls using Wave 13 data (Nov–Dec 2020);[38] and (iii) computing the preloaded health index for each respondent using Wave 14 data (Nov–Dec 2021).

Respondents were shown their index and its components, asked to rate agreement (1 = strongly disagree, 5 = strongly agree), and allowed to adjust if they disagreed (scores 1–3). Most (80 percent) agreed and made no changes. Figure 9 shows the pre-computed and final distribution.

Table 12 Items of the Health Core Module of the LISS Panel Used to Construct the Health Index

Question LISS	*Description*	*Weight*
ch20m020	Did your physical health or emotional problems hinder your daily activities over the past month, for instance in going for a walk, walking up stairs, dressing yourself, washing yourself, visiting the toilet?	1.40
ch20m022	Did your physical health or emotional problems hinder your work/study over the past month, for instance in your job, the housekeeping, taking care of the children, doing volunteer work, or in school?	1.22
ch20m025	Getting up from a chair in which you sat for some time	0.41
ch20m028	Crouching, kneeling, crawling on all fours	0.65
ch20m015	How often did you feel happy during the last month	2.34

[37] Sub-dimensions were selected by reviewing the SF-12 health survey and prioritizing similar objective information available in our data. We also excluded dimensions with little variation to increase distributional variance.

[38] Regression coefficients were used as relative weights; final weights appear in Table 12.

Question LISS	*Description*	*Weight*
ch20m080-097	Has a physician told you this last year that you suffer from one of the following diseases/problems? [list of health problems]	0.45
ch20m169-183	Are you currently taking medicine at least once a week for [list of diseases]	0.73
ch20m018	Do you suffer from any kind of long-standing disease, affliction or handicap, or do you suffer from the consequences of an accident?	0.46
ch20m233	How is your eyesight with (reading) glasses, computer glasses or contact lenses?	1.46
ch20m070-078	Do you regularly suffer from [list of health problems]	0.88

D.2.2 Social relations Index

The Social Relations Index was constructed similarly, using data from the Social Integration and Leisure Core Module. It includes 10 sub-dimensions (Table 13).[39]

The index ranges from 0 to 100. Steps: (i) rescale sub-dimensions to 0–10; (ii) estimate weights via regression of satisfaction with social contacts (1–10 scale) on the 10 sub-dimensions and demographics using Wave 13 data (Oct–Nov 2020);[40] and (iii) compute the index using Wave 14 data (Oct–Nov 2021).

Respondents were shown their index and its components, asked to rate agreement (1 = strongly disagree, 5 = strongly agree), and allowed to adjust if they disagreed (scores 1–3). Most (88 percent) agreed and made no changes. Figure 10 shows the pre-computed and final distribution.

Table 13 Items of the Social Relations and Leisure Core Study Used to Construct the Social Relations Index

Question LISS	*Description*	*Weight*
cs20m288	Missing having people around.	2.74
cs20m285	There are enough people I can count on in case of a misfortune.	0.97
cs20m287	There are enough people to whom I feel closely connected.	1.93

[39] Dimensions were chosen based on literature on social relationship measures and available LISS data, excluding dimensions with little variation.

[40] Final weights appear in Table 13.

Question LISS	*Description*	*Weight*
cs20m290	Spend an evening with family.	0.73
cs20m291	Spend an evening with someone from the neighborhood.	1.23
cs20m292	Spend an evening with someone outside your neighborhood.	0.78
cs20m003-057	Participation in social organizations	0.24
aantalhh & woonvorm	Living alone or with other people	0.24
belbezig	Currently working or studying.	0.72
cs20m436	Use of social media.	0.41

E Sensitivity Analyses

Appendix E.1 reproduces all main results using the unweighted baseline sample (N = 2,050). Appendix E.2 uses the unweighted full sample ($N = 2,186$), including 136 respondents excluded in the baseline sample for speeding; see Appendix D.1 for sample selection. All findings are robust.

E.1 Unweighted Baseline Sample

E.1.1 Prevalence of Incomplete Preferences

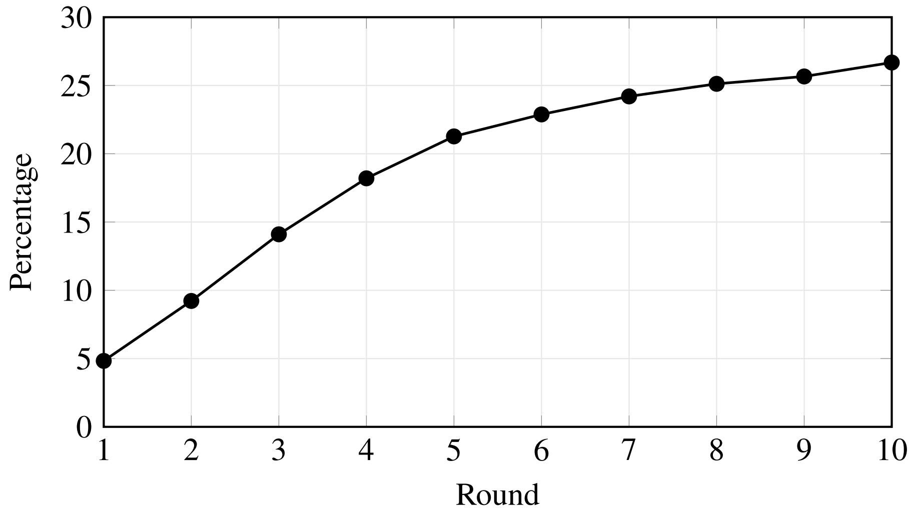

Figure 16 Cumulative prevalence of "Don't know" responses by elicitation round
Notes: Unweighted baseline sample (N = 2,050).

E.1.2 Distribution of Well-Being Dimensions and Adjustments

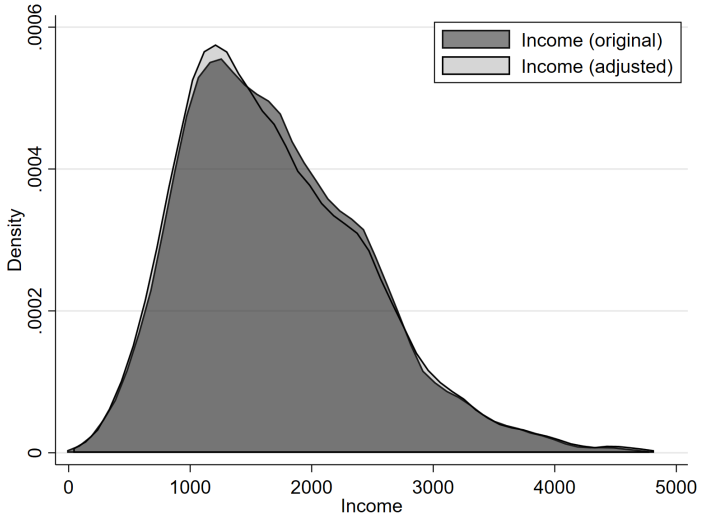

Figure 17 Original (Pre-loaded) and Adjusted Distribution of Income
Notes: Unweighted baseline sample (N = 2,050).

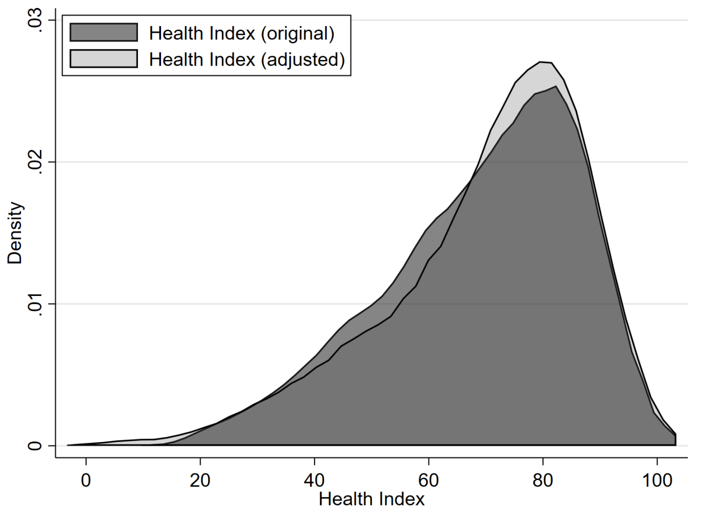

Figure 18 Original (Pre-loaded) and Adjusted Distribution of Health Index
Notes: Unweighted baseline sample (N = 2,050).

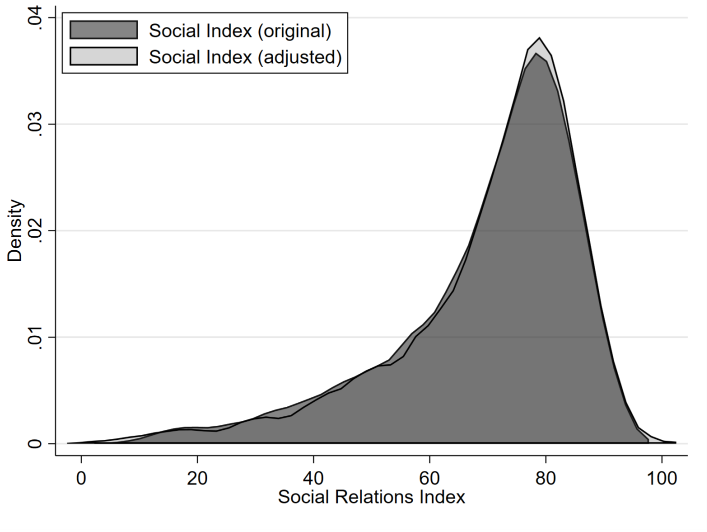

Figure 19 Original (Pre-loaded) and Adjusted Distribution of Social Relation Index
Notes: Unweighted baseline sample (*N* = 2,050).

E.1.3 Determinants of Incompleteness

Table 14 Determinants of incompleteness: robustness across specifications, unweighted sample

Model	(1)	(1')	(2)	(2')	(3)	(3')	(4)	(4')
Dep. Variable	"Don't know"				$\Delta y^{\star}$		"Incompleteness degree"	
Logarithm income	0.107	0.108	0.181	0.194	0.0112	0.0240	-0.274***	-0.290***
	(0.0740)	(0.0708)	(0.133)	(0.126)	(0.00776)	(0.0164)	(0.0665)	(0.0625)
Health index	-0.00572**	-0.00626**	-0.0102**	-0.0113**	-0.000528*	-0.00142**	-0.00598***	-0.00647***
	(0.00244)	(0.00246)	(0.00440)	(0.00447)	(0.000283)	(0.000578)	(0.00209)	(0.00209)
Social relations index	-0.00130	-0.00184	-0.00192	-0.00270	-0.000345	-0.000462	-0.00169	-0.00235
	(0.00256)	(0.00253)	(0.00454)	(0.00448)	(0.000278)	(0.000591)	(0.00205)	(0.00206)
Female	-0.192**	-0.192**	-0.321**	-0.335**	-0.00962	-0.0413**	-0.141**	-0.143**
	(0.0791)	(0.0791)	(0.143)	(0.143)	(0.00815)	(0.0174)	(0.0687)	(0.0686)
Above Median Age	-0.333***		-0.606***		-0.0486***		-0.353***	
	(0.104)		(0.189)		(0.0104)		(0.0877)	
Age		-0.0181***		-0.0329***		-0.00439***		-0.0184***
		(0.00303)		(0.00548)		(0.000701)		(0.00267)
University degree	0.208*		0.377*		-0.00194		0.122	
	(0.109)		(0.194)		(0.0109)		(0.0948)	
Pensioner	-0.247**		-0.397*		-0.0146		-0.242**	
	(0.116)		(0.214)		(0.0125)		(0.100)	
Occupation: paid employee = 1	-0.0556		-0.0495		0.00544		-0.0348	
	(0.104)		(0.189)		(0.0100)		(0.0856)	
Occupation: unemployed = 1		0.0897		0.174		0.0171		-0.115
		(0.239)		(0.437)		(0.0577)		(0.187)
Openness	-0.0236	0.00121	-0.0307	0.0136	0.0105	0.00247	0.0434	0.0507
	(0.0824)	(0.0812)	(0.148)	(0.145)	(0.00852)	(0.0183)	(0.0714)	(0.0707)
Neurotic	0.0296	0.00773	0.0630	0.0219	0.00869	-0.000803	0.00947	-0.00932
	(0.0634)	(0.0635)	(0.115)	(0.116)	(0.00633)	(0.0146)	(0.0532)	(0.0538)
Conscientiousness	-0.178*	-0.169*	-0.315*	-0.286	-0.00723	-0.0380*	-0.119	-0.106
	(0.0996)	(0.0989)	(0.180)	(0.178)	(0.0102)	(0.0227)	(0.0850)	(0.0850)
Agreeableness	-0.0970	-0.0975	-0.188	-0.192	-0.0102	-0.0233	-0.0983	-0.0969
	(0.0811)	(0.0810)	(0.148)	(0.146)	(0.00809)	(0.0183)	(0.0674)	(0.0676)
Extraversion	-0.0408	-0.0392	-0.0829	-0.0810	-0.00146	-0.0106	-0.0384	-0.0391
	(0.0599)	(0.0591)	(0.110)	(0.107)	(0.00572)	(0.0137)	(0.0513)	(0.0508)
Did the questions make you think?	0.323***	0.310***	0.565***	0.536**	0.0112	0.0581***	0.211**	0.196**
	(0.118)	(0.118)	(0.219)	(0.220)	(0.0107)	(0.0224)	(0.0941)	(0.0943)
Did you find the questions unclear?	0.487***	0.513***	0.857***	0.911***	0.0525***	0.115***	0.396***	0.424***
	(0.139)	(0.139)	(0.242)	(0.244)	(0.0198)	(0.0353)	(0.120)	(0.120)
Did you find the questions interesting?	-0.0410	-0.0551	-0.0386	-0.0574	-0.0153	-0.000878	-0.0464	-0.0650
	(0.144)	(0.145)	(0.268)	(0.270)	(0.0153)	(0.0294)	(0.118)	(0.119)
Time last round	0.165***	0.166***	0.293***	0.296***	0.0283***	0.0463***	0.158***	0.160***
	(0.00950)	(0.00957)	(0.0190)	(0.0194)	(0.00140)	(0.00146)	(0.00659)	(0.00671)
Threshold 1							-1.542***	-2.509***
							(0.539)	(0.551)
Threshold 2							-1.340**	-2.307***
							(0.533)	(0.544)
Threshold 3							-1.275**	-2.242***
							(0.532)	(0.544)
Threshold 4							-1.207**	-2.174***
							(0.532)	(0.544)
Threshold 5							-1.125**	-2.092***
							(0.532)	(0.544)
Threshold 6							-0.952*	-1.919***
							(0.533)	(0.545)
Threshold 7							-0.674	-1.640***
							(0.533)	(0.545)
Threshold 8							-0.242	-1.206**
							(0.534)	(0.547)
Threshold 9							0.412	-0.550
							(0.538)	(0.548)
Constant	-1.285**	-0.486	-2.196**	-0.810	-0.0885	0.287**		
	(0.618)	(0.631)	(1.098)	(1.118)	(0.0657)	(0.146)		
Estimation method	Probit	Probit	Logit	Logit	OLS	OLS	Oprobit	Oprobit
N	1,953	1,953	1,953	1,953	1,953	1,953	1,953	1,953
(Pseudo) R^2	0.309	0.309	0.308	0.309	0.474	0.352	0.191	0.191

Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Notes: In models (4) and (4'), the dependent variable "Incompleteness degree" ranges from 0 to 9 and equals the maximum possible rounds minus the rounds answered. A value of 0 means all rounds were answered; a value of 9 means the respondent answered 'I don't know' in the first round.

E.1.4 Distribution of Hurwicz equivalent incomes

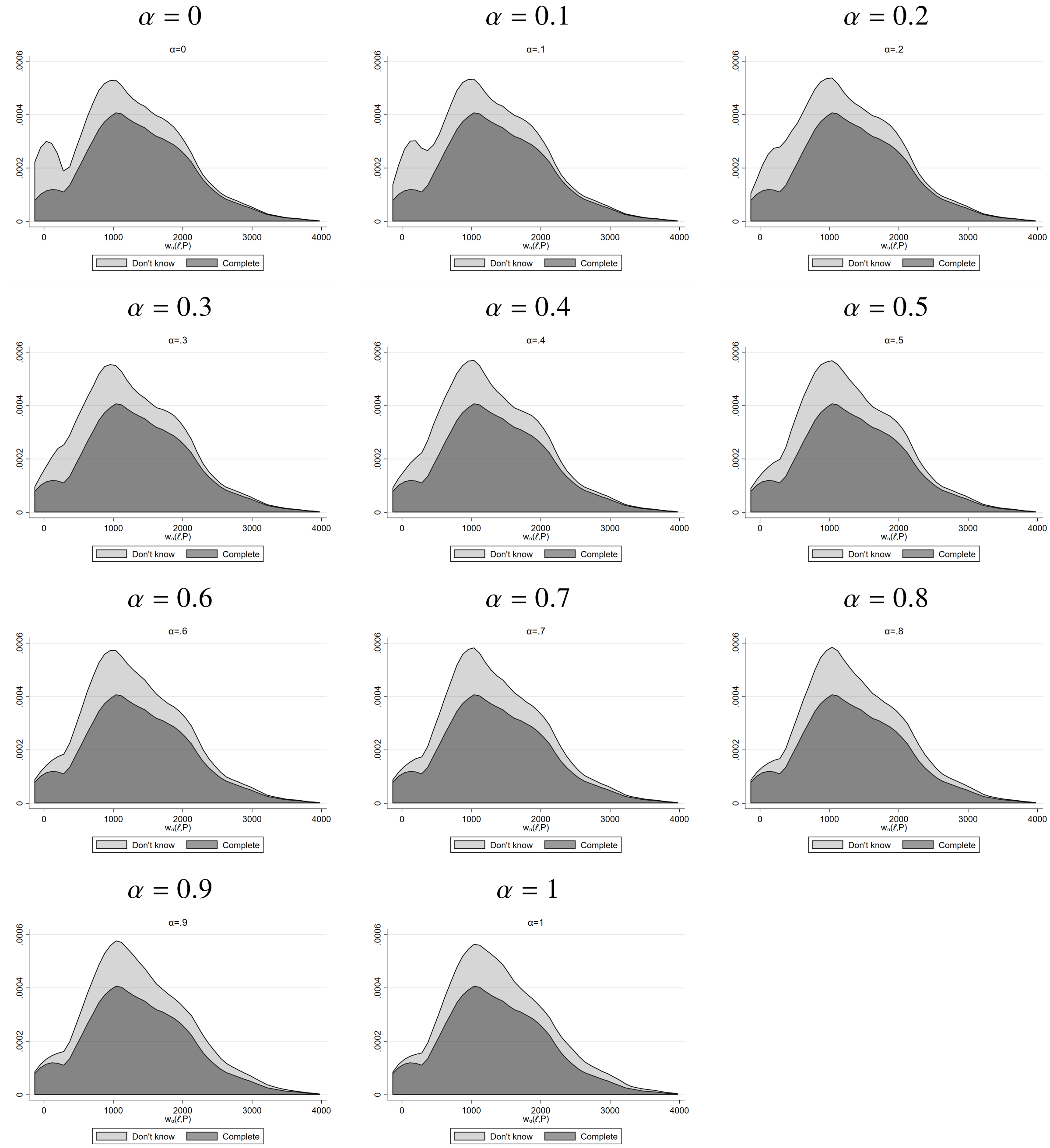

Figure 20 Hurwicz equivalent incomes across α values
Notes: Kernel density of Hurwicz equivalent incomes by α. Epanechnikov kernel, bandwidth 138.5. Unweighted baseline sample (N = 2,050).

E.1.5 Characteristics of the Worst-Off

Table 15 Characteristics of the Worst-Off

	$\alpha = 0$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$
Income	1653	1272	1133	1119	1100
Health	58	58	58	58	59
Social relations	64	63	64	63	64
Age	59	56	54	54	54
Female (%)	56.6	60.0	59.5	61.0	60.5
Married (%)	64.9	68.3	68.8	68.3	67.3
Unemployed (%)	3.4	5.9	6.8	7.3	6.8
Pensioner (%)	30.2	19.5	16.1	16.1	16.1
University degree (%)	9.8	9.3	9.8	10.2	10.2
Non-Dutch background (%)	21.7	26.6	25.2	25.2	25.2
Don't know (%)	68.3	46.3	31.7	26.3	20.0
Rounds	4.2	6.3	7.6	8.0	8.5
N	205	205	205	205	205

Notes: Each column reports averages for the bottom decile under the Hurwicz equivalent-income for a given α. Unweighted baseline sample (N = 2,050).

E.1.6 S-Gini Inequality Indices Across α Values

Table 16 S-Gini Inequality Indices Across α Values

	Round									
α	1	2	3	4	5	6	7	8	9	10
	Bottom-Sensitivity Parameter: $\delta = 2$ (Gini)									
0	42.4	40.0	39.4	39.2	39.2	39.1	39.1	39.1	39.1	39.1
0.25	36.0	36.3	36.5	36.5	36.5	36.5	36.5	36.5	36.5	36.5
0.5	33.0	34.1	34.5	34.6	34.7	34.8	34.8	34.8	34.8	34.8
0.75	31.5	32.8	33.4	33.7	33.8	33.9	33.9	33.9	33.9	33.9
1	30.7	32.1	32.9	33.3	33.5	33.6	33.6	33.7	33.7	33.7
	Bottom-Sensitivity Parameter: $\delta = 5$									
0	42.4	39.9	39.4	39.2	39.1	39.1	39.1	39.1	39.1	39.1
0.25	36.0	36.3	36.5	36.5	36.5	36.5	36.5	36.5	36.5	36.5
0.5	33.0	34.1	34.5	34.6	34.7	34.8	34.8	34.8	34.8	34.8
0.75	31.5	32.8	33.4	33.7	33.8	33.9	33.9	33.9	33.9	33.9
1	30.7	32.1	32.9	33.3	33.5	33.6	33.6	33.7	33.7	33.7
	Bottom-Sensitivity Parameter: $\delta = 5$									
0	77.2	73.1	72.1	71.7	71.5	71.5	71.5	71.5	71.5	71.5
0.25	64.6	65.3	65.6	65.8	65.8	65.9	65.9	65.9	65.9	65.9
0.5	58.3	60.9	61.9	62.3	62.5	62.7	62.7	62.7	62.8	62.8
0.75	55.1	58.4	60.0	60.7	61.0	61.2	61.3	61.4	61.4	61.4
1	53.4	56.9	58.9	59.9	60.5	60.7	60.9	60.9	61.0	61.0
	Bottom-Sensitivity Parameter: $\delta = 10$									
0	82.4	85.9	85.8	85.6	85.6	85.5	85.5	85.5	85.5	85.5
0.25	78.0	79.8	80.6	80.9	81.1	81.2	81.3	81.3	81.3	81.3
0.5	70.5	74.6	76.4	77.2	77.6	77.8	77.9	78.0	78.0	78.0
0.75	66.7	71.5	74.2	75.5	76.1	76.5	76.6	76.7	76.7	76.7
1	64.6	69.5	72.8	74.5	75.5	75.9	76.2	76.3	76.3	76.3

Notes: S-Gini indices of Hurwicz equivalent incomes by α and δ. Columns indicate ABDC elicitation round. Unweighted baseline sample (N = 2,050).

E.2 Unweighted Full Sample

E.2.1 Prevalence of Incomplete Preferences

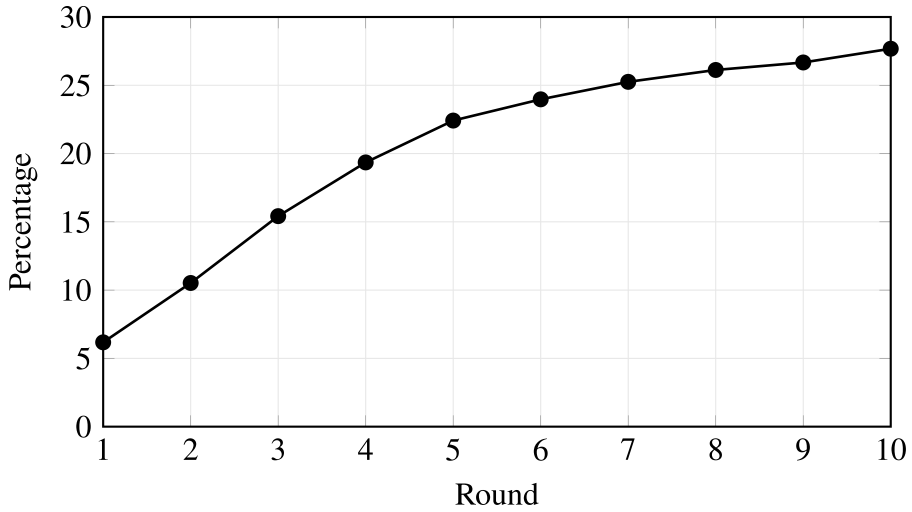

Figure 21 Cumulative prevalence of "Don't know" responses by elicitation round
Notes: Unweighted full sample (N = 2,186).

E.2.2 Distribution of Well-Being Dimensions and Adjustments

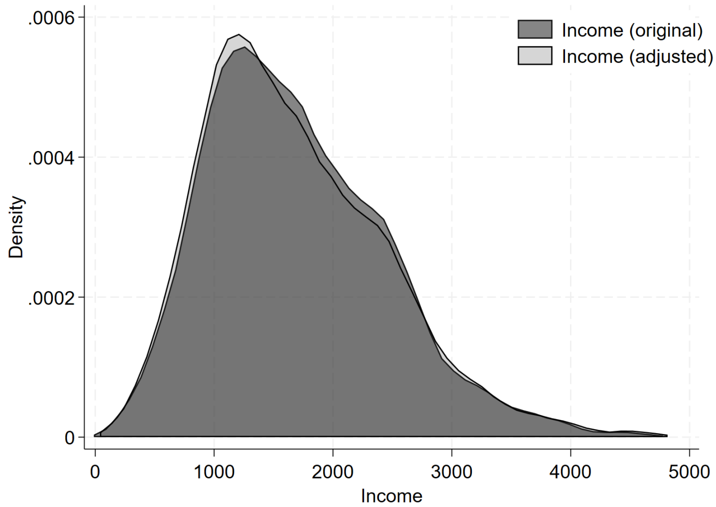

Figure 22 Original (Pre-loaded) and Adjusted Distribution of Income
Notes: Unweighted full sample (*N* = 2,186).

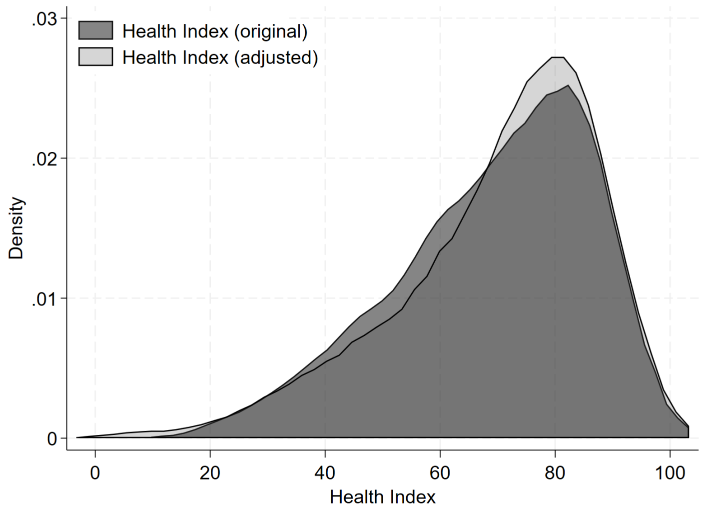

Figure 23 Original (Pre-loaded) and Adjusted Distribution of Health Index
Notes: Unweighted full sample (*N* = 2,186).

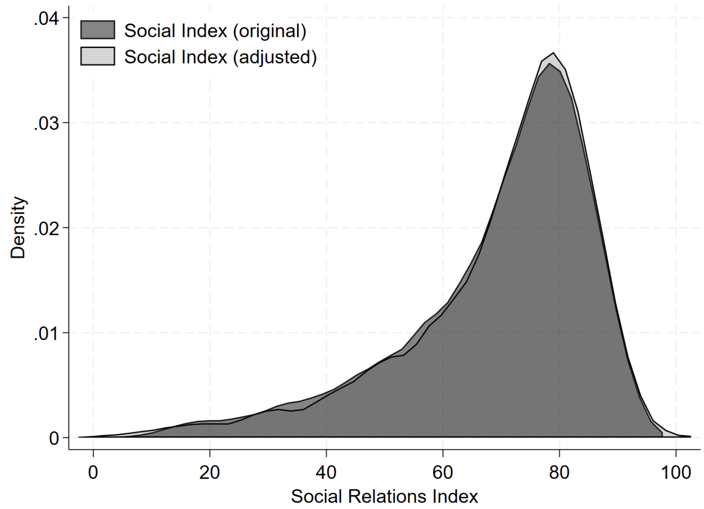

Figure 24 Original (Pre-loaded) and Adjusted Distribution of Social Relations Index
Notes: Unweighted full sample (N = 2,186).

E.2.3 Determinants of Incompleteness

Table 17 Determinants of incompleteness: robustness across specifications. Extended Sample

Model	(1)	(1')	(2)	(2')	(3)	(3')	(4)	(4')
Dep. Variable	"Don't know"				$\Delta y^{\star}$		"Incompleteness degree"	
Logarithm income	0.0326	0.0265	0.0522	0.0569	-0.00563	0.00619	-0.331***	-0.349***
	(0.0712)	(0.0687)	(0.128)	(0.122)	(0.0106)	(0.0167)	(0.0677)	(0.0643)
Health index	-0.00687***	-0.00745***	-0.0117***	-0.0129***	-0.000773**	-0.00173***	-0.00671***	-0.00732***
	(0.00235)	(0.00236)	(0.00419)	(0.00425)	(0.000353)	(0.000586)	(0.00206)	(0.00206)
Social relations index	-0.00197	-0.00275	-0.00315	-0.00444	-0.000668**	-0.000709	-0.00300	-0.00385*
	(0.00249)	(0.00246)	(0.00438)	(0.00434)	(0.000329)	(0.000607)	(0.00201)	(0.00203)
Female	-0.155**	-0.156**	-0.254*	-0.267**	-0.00563	-0.0348**	-0.106	-0.112*
	(0.0763)	(0.0757)	(0.137)	(0.136)	(0.00997)	(0.0174)	(0.0678)	(0.0668)
Above Median Age	-0.313***		-0.571***		-0.0573***		-0.340***	
	(0.100)		(0.181)		(0.0124)		(0.0851)	
Age		-0.0184***		-0.0335***		-0.00464***		-0.0190***
		(0.00290)		(0.00522)		(0.000691)		(0.00262)
University degree	0.249**		0.451**		0.00708		0.160*	
	(0.104)		(0.183)		(0.0126)		(0.0901)	
Pensioner	-0.297***		-0.484**		-0.0246*		-0.265***	
	(0.112)		(0.204)		(0.0142)		(0.0963)	
Occupation: paid employee = 1	-0.0780		-0.0886		-0.00537		-0.0519	
	(0.0963)		(0.172)		(0.0124)		(0.0813)	
Occupation: unemployed = 1		-0.0978		-0.174		-0.0242		-0.277
		(0.238)		(0.437)		(0.0547)		(0.188)
Openness	-0.0787	-0.0499	-0.131	-0.0817	-0.00168	-0.0107	-0.0294	-0.0203
	(0.0800)	(0.0790)	(0.142)	(0.139)	(0.00961)	(0.0184)	(0.0684)	(0.0676)
Neurotic	0.0211	0.00112	0.0491	0.0105	0.00681	-0.00233	-0.00172	-0.0205
	(0.0601)	(0.0600)	(0.107)	(0.108)	(0.00737)	(0.0144)	(0.0514)	(0.0517)
Conscientiousness	-0.218**	-0.205**	-0.390**	-0.352**	-0.0238**	-0.0488**	-0.177**	-0.161**
	(0.0954)	(0.0948)	(0.170)	(0.168)	(0.0114)	(0.0225)	(0.0813)	(0.0814)
Agreeableness	-0.193**	-0.187**	-0.359**	-0.351**	-0.0313***	-0.0451**	-0.199***	-0.191***
	(0.0778)	(0.0775)	(0.140)	(0.139)	(0.00982)	(0.0183)	(0.0655)	(0.0652)
Extraversion	-0.00953	-0.00822	-0.0243	-0.0251	0.00602	-0.00384	-0.00313	-0.00450
	(0.0577)	(0.0569)	(0.104)	(0.102)	(0.00636)	(0.0134)	(0.0489)	(0.0485)
Did the questions make you think?	0.334***	0.324***	0.579***	0.556***	0.0157	0.0613***	0.219**	0.207**
	(0.111)	(0.111)	(0.202)	(0.203)	(0.0125)	(0.0222)	(0.0907)	(0.0907)
Did you find the questions unclear?	0.499***	0.513***	0.867***	0.905***	0.0582**	0.125***	0.398***	0.418***
	(0.131)	(0.131)	(0.225)	(0.226)	(0.0236)	(0.0355)	(0.117)	(0.116)
Did you find the questions interesting?	-0.216	-0.229*	-0.350	-0.366	-0.0545***	-0.0421	-0.260**	-0.274**
	(0.132)	(0.133)	(0.237)	(0.240)	(0.0204)	(0.0303)	(0.115)	(0.115)
Time last round	0.164***	0.166***	0.289***	0.293***	0.0281***	0.0468***	0.148***	0.151***
	(0.00910)	(0.00920)	(0.0180)	(0.0185)	(0.00144)	(0.00151)	(0.00635)	(0.00644)
Threshold 1							-2.933***	-3.923***
							(0.568)	(0.567)
Threshold 2							-2.743***	-3.733***
							(0.564)	(0.564)
Threshold 3							-2.683***	-3.673***
							(0.564)	(0.564)
Threshold 4							-2.622***	-3.612***
							(0.564)	(0.564)
Threshold 5							-2.547***	-3.537***
							(0.564)	(0.564)
Threshold 6							-2.389***	-3.379***
							(0.566)	(0.565)
Threshold 7							-2.150***	-3.138***
							(0.567)	(0.567)
Threshold 8							-1.783***	-2.766***
							(0.569)	(0.569)
Threshold 9							-1.302**	-2.279***
							(0.575)	(0.574)
Constant	0.0483	0.881	0.133	1.551	0.253**	0.623***		
	(0.598)	(0.605)	(1.062)	(1.065)	(0.101)	(0.152)		
Estimation method	Probit	Probit	Logit	Logit	OLS	OLS	Oprobit	Oprobit
N	2,073	2,073	2,073	2,073	2,073	2,073	2,073	2,073
(Pseudo) R^2	0.294	0.294	0.293	0.294	0.386	0.335	0.171	0.172

Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Notes: In models (4) and (4'), the dependent variable "Incompleteness degree" ranges from 0 to 9 and equals the maximum possible rounds minus the rounds answered. A value of 0 means all rounds were answered; a value of 9 means the respondent answered 'I don't know' in the first round.

E.2.4 Distribution of Hurwicz equivalent incomes

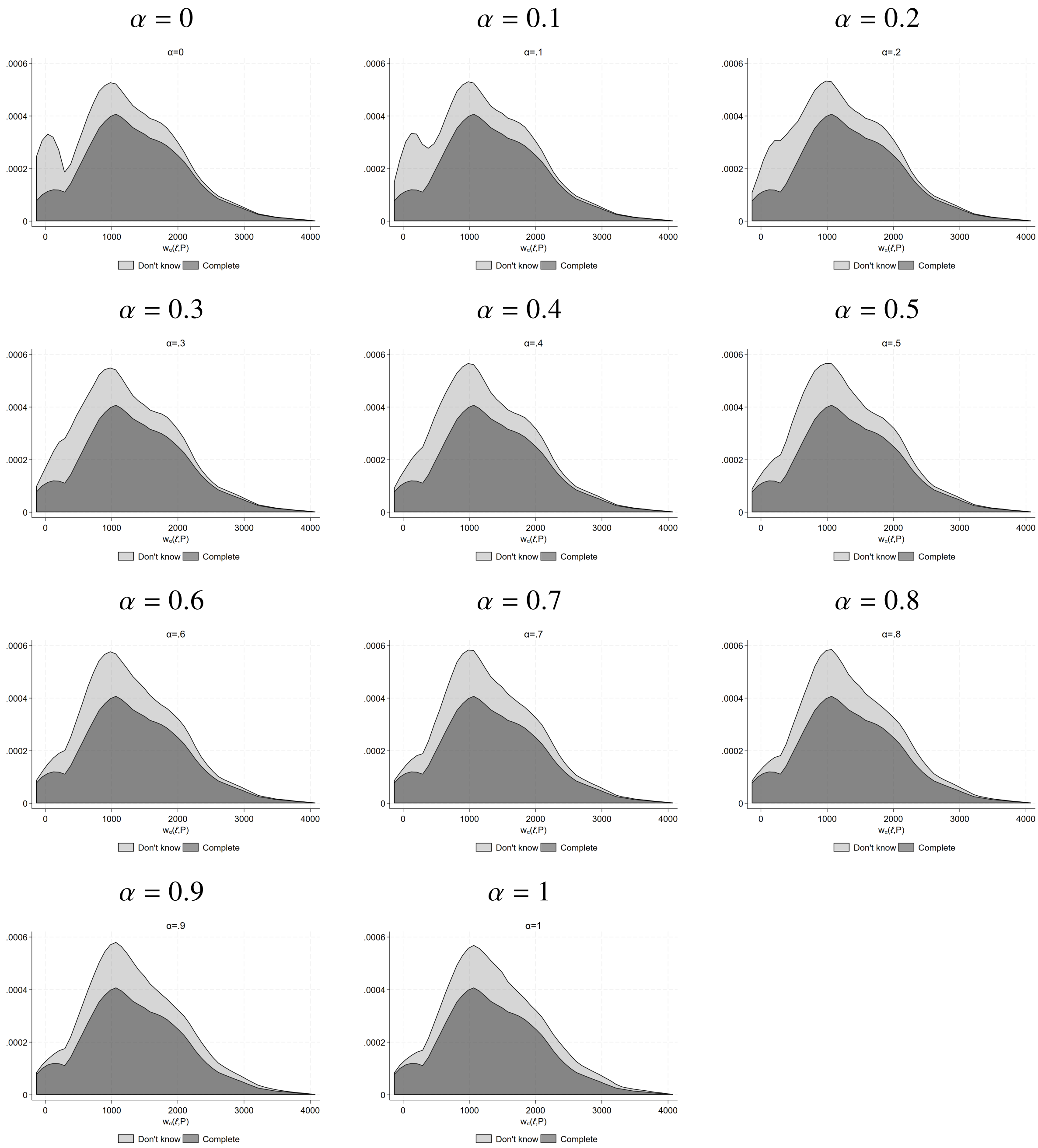

Figure 25 Hurwicz equivalent incomes across α values

Notes: Kernel density of Hurwicz equivalent incomes by α. Epanechnikov kernel, bandwidth 138.5. Unweighted full sample (N = 2,186)

.

E.2.5 Characteristics of the Worst-Off

Table 18 Characteristics of the Worst-Off

	$\alpha = 0$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$
Income	1613	1217	1079	1059	1059
Health	58	58	59	59	59
Social relations	63	62	64	63	63
Age	58	55	53	53	53
Female (%)	56.9	59.2	60.1	60.6	59.6
Married (%)	65.1	68.4	67.4	68.4	67.0
Unemployed (%)	2.8	5.1	6.0	6.9	6.4
Pensioner (%)	28.0	18.4	15.6	14.2	15.6
University degree (%)	11.5	10.1	11.0	11.0	10.6
Non-Dutch background (%)	24.8	28.2	26.8	28.2	29.1
Don't know (%)	71.6	47.3	34.4	29.8	24.8
Rounds	3.9	6.2	7.3	7.7	8.1
N	218	218	218	218	218

Notes: Each column reports averages for the bottom decile under the Hurwicz equivalent-income measure for a given α. Unweighted full sample (N = 2,186).

E.2.6 S-Gini Inequality Indices Across α Values

Table 19 S-Gini Inequality Indices Across α Values

	Round									
α	1	2	3	4	5	6	7	8	9	10
	Bottom-Sensitivity Parameter: $\delta = 2$ (Gini)									
0	43.1	40.8	40.3	40.1	40.0	40.0	40.0	40.0	40.0	40.0
0.25	36.4	36.8	37.0	37.0	37.1	37.1	37.1	37.1	37.1	37.1
0.5	33.3	34.4	34.8	35.0	35.1	35.1	35.1	35.1	35.1	35.1
0.75	31.7	33.0	33.6	33.9	34.0	34.1	34.1	34.1	34.1	34.1
1	30.9	32.3	33.0	33.4	33.6	33.7	33.8	33.8	33.8	33.8
	Bottom-Sensitivity Parameter: $\delta = 5$									
0	78.1	74.5	73.6	73.2	73.1	73.1	73.1	73.1	73.0	73.0
0.25	65.3	66.2	66.6	66.8	66.9	66.9	67.0	67.0	67.0	67.0
0.5	58.7	61.4	62.5	62.9	63.2	63.3	63.3	63.4	63.4	63.4
0.75	55.5	58.8	60.4	61.0	61.4	61.6	61.7	61.7	61.8	61.8
1	53.8	57.2	59.2	60.2	60.7	61.0	61.1	61.2	61.2	61.2
	Bottom-Sensitivity Parameter: $\delta = 10$									
0	81.6	86.1	86.2	86.2	86.2	86.2	86.1	86.2	86.2	86.2
0.25	78.5	80.6	81.5	81.8	82.0	82.1	82.2	82.2	82.2	82.2
0.5	70.9	75.1	76.9	77.7	78.2	78.4	78.5	78.5	78.6	78.5
0.75	67.1	71.9	74.5	75.8	76.5	76.8	77.0	77.0	77.1	77.1
1	65.1	69.9	73.1	74.8	75.7	76.1	76.4	76.5	76.5	76.5

Notes: S-Gini indices of Hurwicz equivalent incomes by α and δ. Columns indicate ABDC elicitation round. Unweighted full sample (N = 2,186).