

Alós-Ferrer, Carlos; Garagnani, Michele

**Working Paper**

## Who likes it more? Using response times to elicit group preferences in surveys

Working Paper, No. 422

**Provided in Cooperation with:**

Department of Economics, University of Zurich

*Suggested Citation:* Alós-Ferrer, Carlos; Garagnani, Michele (2022) : Who likes it more? Using response times to elicit group preferences in surveys, Working Paper, No. 422, University of Zurich, Department of Economics, Zurich, <https://doi.org/10.5167/uzh-224551>

This Version is available at:

<https://hdl.handle.net/10419/267216>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



**University of  
Zurich** <sup>UZH</sup>

University of Zurich  
Department of Economics

Working Paper Series

ISSN 1664-7041 (print)  
ISSN 1664-705X (online)

---

Working Paper No. 422

**Who Likes It More?  
Using Response Times To Elicit Group Preferences  
in Surveys**

Carlos Alós-Ferrer and Michele Garagnani

December 2022

---

# Who Likes It More?

## Using Response Times To Elicit Group Preferences in Surveys

Carlos Alós-Ferrer\*   Michele Garagnani  
University of Zurich   Bocconi University

This Version: December 2022

### Abstract

Surveys and opinion polls are essential instruments to elicit societal preferences and uncover differences between socioeconomic or demographic groups. However, survey data is noisy, and survey bias is ubiquitous, limiting the reliability and usefulness of standard analyses. We provide a new method that uncovers group preferences and unambiguously ranks the relative strength of preference between groups of agents, leveraging the information contained in response times. The method delivers a nonparametric criterion to determine whether a group (defined, e.g., by gender, age cohort, socioeconomic status, political orientation, etc.) prefers an option over its alternative, and whether it does so more strongly than another group, without any assumptions on the underlying noise. We demonstrate the practical value of this method by studying preferences over important socioeconomic topics in a representative sample of the U.K. population. We find that the new method often provides results when tests based on choice frequencies are inconclusive, and also identifies cases where tests are significant but inferences on preferences are unwarranted.

**JEL Classification:** C83 · D11 · D87 · D91

**Keywords:** Survey Data · Revealed Preference · Response Times · Stochastic Choice

## 1 Introduction

Survey data is extensively used to elicit societal preferences and attitudes, ranging from the support for redistributive policies to the willingness to pay for a new product. The advent of online survey platforms and the increased availability of extensive, well-maintained national-level panels have considerably increased the use of survey data in economics, political science, marketing, health research, and other disciplines. For

---

\*Corresponding author: carlos.alos-ferrer@econ.uzh.ch. Zurich Center for Neuroeconomics (ZNE), Department of Economics, University of Zurich. Blümlisalpstrasse 10, 8006 Zurich, Switzerland.

example, surveys and opinion polls are used to gauge whether a given economic policy enjoys sufficient support in the population and is hence politically-sustainable. This is often closely-linked to the estimation of economic preferences in groups. In particular, risk, time, and social preferences determine the support for social security policies (e.g., health or unemployment protection), investments with delayed benefits (e.g., energy policy, infrastructures), and policies affecting future generations (e.g., climate protection) or non-citizens (e.g., migration), respectively (e.g., Alesina and Fuchs-Schündeln, 2007; Falk et al., 2018; Bechtel and Liesch, 2020; Bechtel et al., 2020; Enke et al., 2022). As another example, marketing research relies on surveys to estimate the market demand for new products, often feeding back into the design of the product’s features (e.g., Day, 1968; Vakratsas and Ambler, 1999). An important question of interest in many empirical studies using surveys is how pre-defined groups differ in their attitudes, e.g. whether a redistributive policy would receive stronger support from the left or the right of the political spectrum, whether females would support a policy more than males, whether preferences differ across racial groups, or whether people with a higher socioeconomic status would demand a new service more than others (e.g., Fisman et al., 2006, 2008; Leeper et al., 2020; Giglio et al., 2021; Snowberg and Yariv, 2021; Ortoleva et al., 2021).

Regrettably, survey data is notoriously imprecise. First, even under ideal conditions, human choices are inherently noisy. Extensive evidence shows that individuals very often make different choices when confronted with the same set of options repeatedly (e.g., Tversky, 1969; Camerer, 1989; Hey and Orme, 1994; Alós-Ferrer et al., 2021). This is true even when choices are incentivized and made in a highly-controlled lab environment. The problem is exacerbated for survey data, due e.g. to increased measurement errors or possible lack of respondents’ attention or understanding. A second and even more important problem is that surveys are often biased and responses might not reflect actual preferences. In particular, questions on sensitive topics are often subject to systematic survey misreporting due to *social desirability bias*, i.e. the misreporting of own preferences in surveys to be more aligned with attitudes perceived as socially acceptable (Zaller and Feldman, 1992). Bertrand and Mullainathan (2001) point out that measurement errors in surveys often correlate with individual characteristics and might hide biases of this type, e.g. a reluctance by specific groups to express an unpopular opinion. For example, using data from 20 years of after-election phone surveys covering 184 referenda in Switzerland, Funk (2016) reports sizeable differences between surveys and actual election results and identifies a “liberal bias” where initiatives perceived to be more prosocial or liberal receive a larger stated support in Swiss surveys than in the actual votes. A different but related problem is *experimenter demand*, which affects economic experiments and surveys (Zizzo, 2010). Both survey bias and experimenter demand are important causes of concern which have motivated an extensive literature and increasingly-sophisticated methods to explore the robustness of conclusions derived from self-reported data obtained in surveys or experiments (e.g., Luce and Tukey, 1964; Hainmueller et al., 2015; de Quidt et al., 2018; Gillen et al., 2019).

In this contribution, we develop and illustrate a new method for preference revelation with survey data which leverages a readily-available, additional source of information to improve over and complement existing analyses. Specifically, we rely on the joint use of choice frequencies and response times. The latter are nowadays easy to collect in online platforms and surveys, and indeed are typically already being passively collected (and then neglected) by standard software. Hence, the method we propose adds no cost to existing survey designs, and can be easily implemented whenever survey data (or laboratory data) is collected electronically.

Our method is based on key insights from the cognitive sciences linking response times to strength of preference (e.g. Dashiell, 1937; Moyer and Landauer, 1967; Laming, 1985; Shadlen and Kiani, 2013, among many others), which are receiving increasing attention in economics (e.g. Fudenberg et al., 2018; Baldassi et al., 2020), and which imply that the distribution of response times contains information on the underlying distribution of behavioral noise. Recently, Alós-Ferrer et al. (2021) built upon those regularities to provide individual preference revelation conditions using response times for settings where a single decision maker has repeatedly made the same decision, over and over again.<sup>1</sup> In contrast, our method considers groups of many individuals but requires only one decision per individual, and is hence appropriate to use with electronically-collected survey data as well as laboratory experiments. That is, while Alós-Ferrer et al. (2021) considers the (laboratory) situation where one individual makes a fixed decision many times, here we consider the (survey) case where many respondents make a given decision, but each person makes that decision exactly once. This setting makes our techniques immediately useful for a wider set of applications, including standard survey data.

Our main theoretical results assume a generalization of a utilitarian framework, which is equivalent to a standard population-level interpretation of random utility models as used in economics (McFadden, 1974, 2001; Anderson et al., 1992), marketing (e.g., Baltas and Doyle, 2001; Feng et al., 2022), political science (e.g., Nownes, 1992; Karp, 2009), and many other fields. That is, the preferences of a group are described by a distribution over utility functions, or, equivalently, a utility function and a distribution of behavioral noise, which can be seen as the individual deviations from the utility function (Block and Marschak, 1960). In this context, survey bias is intuitively easy to conceptualize and visualize. Imagine that preferences within a group mainly favor option  $x$  over option  $y$ , but social desirability bias creates a tendency to provide a  $y$  response. Within a random utility model, this will result in an *asymmetric* distribution, where the median differs from the mean. Hence, it is possible that more than half of the surveyed individuals give a  $y$  response (reflecting the median), while in reality the group preference (the mean) is in favor of  $x$ . Analogously, it is possible that the bias affects a group more than another (e.g., due to gender effects or cohort differences), biasing results on group differences.

---

<sup>1</sup>We refer the reader to Alós-Ferrer and Garagnani (2022a,b) and the references therein for additional evidence and discussion on *psychometric* and *chronometric effects*. Alós-Ferrer et al. (2022) extend the results in Alós-Ferrer et al. (2021) to analyze whether apparent transitivity violations are due to actual failures of transitivity or just to behavioral noise.

As pointed out by Alós-Ferrer et al. (2021) (there for the case of individuals), the use of choice frequencies alone might lead to unwarranted conclusions over preferences.

Current standard practice to analyze survey data focuses on statistical tests using choice proportions only. For example, to examine preference differences across groups, the analyst compares the proportion of choices in favor of one option between the two groups. If the two proportions are statistically different, according to, say, a test of proportions, it is concluded that one group prefers the option more than the other, and the researcher has no other instrument to further evaluate comparison across groups. This approach can only reveal differences in stated preferences, which might be heavily biased, and fails to reveal differences in actual preferences.

In the presence of survey bias, the question of preference revelation differs from the question of whether a larger fraction of individuals choose a given response. Thus, when applying our results to a dataset, for instance for differences across groups, four situations might arise. First, it might be that a statistical test using choice proportions fails to detect a significant difference, but our method delivers the conclusion that one group does exhibit a stronger preference. That is, our method is able to get more out of the data. This is possible simply because our method uses response times, which the analysis of choice proportions ignores. In particular, in this case our method might signal a false negative. It might also be the case that a group's preferences differ from those of others but social desirability bias leads group members to conform with socially-acceptable views. For instance, some older people might disagree with vocal statements that richer countries should finance greenhouse reduction policies alone, but might be reluctant to express their opinion. In this case, survey choice proportions will not show significant differences, but our techniques can detect the difference in group preferences.

Second, it might be that a statistical test detects a significant difference, but our method signals that it is unwarranted to conclude that there is a difference in preferences across groups. This is possible because choice proportions might reflect group-specific survey bias. That is, survey bias might affect different groups differently, artificially exacerbating minor or nonexistent differences in preferences. For example, some women might oppose gender quotas, but might feel compelled to express an opinion in their favor.

Third, it might be that a statistical test detects a significant difference and, when response times are taken into account, our method agrees, in the sense that the same group is revealed to exhibit a stronger preference. In this case, our result strengthens the positive conclusion and suggests the absence of strong biases.

Last, it might occur that a statistical test fails to detect a significant difference, and our method also concludes that it is unwarranted to assume a difference in preferences across groups. Strictly speaking, in this case the researcher cannot draw any conclusion, but the failure to detect a preference difference suggests that the lack of significance might not be due to countervailing biases obscuring an actual difference.

To demonstrate and illustrate the practical value of the new method, we collected survey data and response times for a representative sample of the U.K. population (pre-registered), including a wide range of important socio-economic issues, e.g. inequality, gender discrimination, and environmental policies. These questions, and the conclusions we obtain for each of them, are of course of independent interest. Our focus, however, is on showcasing the applicability and implications of the proposed method for a variety of important issues, and hence we designed the survey to explicitly cover different fields of application.

Our data shows that all four cases described above are relevant. We demonstrate that existing procedures might indeed lead to unwarranted (positive or negative) conclusions on fundamental issues in a sizeable proportion of cases. This was the case for almost 40% of the questions we implemented. That is, in these cases, our method and statistical tests relying on choice proportions deliver opposite conclusions. For example, we find a number of cases where our method allows to reveal preference orderings but existing statistical procedures would lead to accept the null hypothesis. Hence, the method can often extract more information from the dataset than statistical tests relying on choice frequencies only. We also find several instances of the opposite situation. In these cases, a researcher using only choice data would conclude that one group prefers an option over the other more than another group, while the data does not actually support this statement once response times are used to reveal preferences. Of course, we also find instances of positive and negative concordance, where our method arrives at the same conclusion as a test based only on choice frequencies.

The results reported in this work are of interest for market research, welfare analysis, normative economics in general, political science, and other fields. For instance, the analysis can potentially identify which group is more receptive to a new product, more likely to endorse a new political candidate, or more likely to support a social change. They can also facilitate the analysis of staircase designs where preferences are elicited through sequences of interrelated questions (e.g. Falk et al., 2018). Also, although the results are formulated in terms of comparisons across groups, obviously they also apply if each “group” is replaced with a single individual and the dataset contains multiple observations of the same choice for that individual, as in the applications discussed and illustrated in Alós-Ferrer et al. (2021).

The remainder of the paper is structured as follows. Section 2 describes our formal population framework and derives our theoretical results. Section 3 describes the survey data collection and presents the results of the empirical application of the method, showcasing a number of situations of interest, e.g. when statistical methods fail to provide a conclusion or lead to unwarranted conclusions. Section 4 explores the limits of the framework by introducing a more demanding concept of *uniform* preference revelation, identifying a necessary and sufficient condition for it, and applying it to the data. Section 5 concludes. The Appendix contains the formal proofs of our results, the detailed transcript of the survey, and a question-by-question analysis of the survey data.

## 2 A Formal Framework for Response Times in Surveys

The framework we work with builds upon the standard additive random utility model (RUM), widely used in applied microeconomics, which is equivalent to stochastic preference models (McFadden, 1974, 2001; Anderson et al., 1992). As a model of stochastic choice for an individual agent, an additive RUM postulates that the agent is endowed with a utility function  $u$  over a feasible set, but is affected by random utility shocks. Thus, given a choice between two alternatives  $x$  and  $y$ , realized utilities are  $u(x) + \varepsilon_x$  and  $u(y) + \varepsilon_y$ , respectively, where  $\varepsilon_x, \varepsilon_y$  are zero-mean random variables. Thus, a RUM generates choice probabilities, with the probability of  $x$  being chosen when  $y$  is also available given by

$$p(x, y) = \text{Prob}(u(x) + \varepsilon_x > u(y) + \varepsilon_y) = \text{Prob}(\varepsilon_x - \varepsilon_y > u(y) - u(x)).$$

where tie-breaking conventions are irrelevant for continuously-distributed errors. Under specific assumptions on the distributions of the error terms, one obtains particular models, as the celebrated logit choice (Luce, 1959) or the classical probit choice (Thurstone, 1927). This general setting has become one of the dominant approaches in economics to model the fact that choice is empirically (and overwhelmingly) observed to be stochastic.

Alós-Ferrer et al. (2021) introduced a more general class of RUM models where error terms are modeled directly for utility differences, i.e. the realized utility difference given a choice  $\{x, y\}$  is  $u(x) - u(y) + \varepsilon_{x,y}$  for a mean-zero random variable  $\varepsilon_{x,y}$  and hence

$$p(x, y) = \text{Prob}(\varepsilon_{x,y} > u(y) - u(x)).$$

This class encompasses additive RUMs, but also trembling-hand models (Loomes et al., 2002) where decisions follow a fixed strict preference but pair-specific errors might always occur.

Building upon insights in response times from psychology and neuroscience, Alós-Ferrer et al. (2021) provided sufficient conditions on the distributions of response times conditional on each possible choice ( $x$  or  $y$  for a given pair  $(x, y)$ ) which ensure that any RUM that fits the data (in terms of choices and response times) reveals a preference for, say,  $x$  over  $y$ , in the sense that  $u(x) > u(y)$  for the underlying  $u$ . Those results are powerful because they guarantee that an option is preferred to another for *any* utility function and *any* distribution of the error term that the analyst might consider, and hence the results are completely non-parametric and independent of functional forms.

An alternative interpretation of (additive) RUMs follows a utilitarian approach (e.g., d'Aspremont and Gevers, 2002). In this approach, the random utility terms are interpreted to model unobserved heterogeneity in a population of agents. That is, instead of considering a distribution of choices for a single agent with trial-to-trial variability, one assumes a distribution of agents, each endowed with a fixed utility drawn from a distribution, such that  $u(x) - u(y)$  is the mean of the distribution of utilities for the particular



choice  $\{x, y\}$ . Thus,  $u$  measures the *utilitarian welfare* and a revealed preference for  $x$  over  $y$  means that  $x$  is to be preferred, in (utilitarian) welfare terms, to  $y$ . Of course, the *caveat* of this approach (and of the utilitarian approach as a whole) is that it requires interpersonal comparability of the units in which utility is measured.

If applied to data of a given individual, the results in Alós-Ferrer et al. (2021) require multiple repetitions of the same choice. The utilitarian interpretation of a RUM opens the door to a different kind of applications. In those, a dataset contains the choices and response times of a population for a fixed choice, but only one choice per individual. This is particularly attractive for field and survey settings, where choice repetition might be difficult to implement, but the collection of large numbers of responses for short questionnaires entails little difficulty.

This section derives new results within the framework described above, which allow for preference revelation in terms of relative strength of preference across different sub-groups of a population. Specifically, our main result identifies a simple, joint condition on choice frequencies and the distributions of response times on the groups which, if fulfilled, allows to unambiguously rank the relative strength of preference of the groups. In other words, the condition guarantees that group  $A$  prefers  $x$  over  $y$  *more* than group  $B$ , for any model that fits the data. The result is nonparametric, because the conclusion follows independently of which utility functions and models of noise are used to fit the data.

In addition to the results on comparisons across groups, we also provide preference revelation results for groups, along the lines of Alós-Ferrer et al. (2021) but for survey data. That is, focusing on a single group, we identify a condition on the choice frequencies and response times of that group alone which, if fulfilled, guarantees that an aggregate preference exists for that group, again independently of which utility and model of noise is used to fit the data.

The main result on the comparison of preferences across groups, however, is independent of whether a preference for  $x$  over  $y$  for the separate groups is actually revealed by the data or not. That is, it is perfectly possible that the data does not allow to conclude that either group prefers  $x$  over  $y$ , and yet the researcher can conclude that the first group favors  $x$  over  $y$  more strongly than the second group. This allows discussing group differences even when within-group preference revelation fails.

Further, it is not necessary that the comparison refers to the same choice pair. In their most general formulation, the results allow to conclude that a certain group prefers  $x$  over  $y$  *more* than another group prefers  $z$  over  $w$ , for any four alternatives  $x, y, z, w$ . This is interesting because, even for identical questions on surveys, it can always be argued that different groups might perceive the options differently. For example, a given level of health care has different consequences for men and women purely on biological grounds, or a policy question might be formulated in terms of instruments (e.g., taxes or subsidies) rather than actual outcomes (e.g., wealth distributions). While this difficulty is usually glossed over on pragmatic grounds, we are able to tackle it head on. The

results can be immediately applied to (i) the comparison of preferences for a fixed pair  $(x, y)$  across two different groups, (ii) the comparison of preferences for group-tailored choice pairs,  $(x_1, y_1)$  for one group and  $(x_2, y_2)$  for another group; (iii) the comparison of preferences across two different pairs  $(x, y)$  and  $(z, w)$  for a fixed group, and (iv) the staircase-like comparison of preferences across  $(x, y)$  and  $(y, z)$  for a fixed group, to mention just the most obvious possibilities.

The conditions identified here are formulated in terms of weakenings of first-order stochastic dominance between appropriate, conditional distributions of response times, or obvious reformulations thereof. For each actual dataset, how weak the weakening is depends on actual choice proportions, so that, generally speaking, a larger percentage of choices in favor of an option requires less information from response times (i.e. the conditions become stronger), and vice versa.

## 2.1 The Formal Setting

Let  $X$  be a finite set of options, and denote by  $C = \{(x, y) \mid x, y \in X, x \neq y\}$  the set of all binary choice problems, so  $(x, y)$  and  $(y, x)$  both represent the problem of choice between  $x$  and  $y$ . Let  $D \subseteq C$  be the set of choice problems on which we have data, assumed to be non-empty and symmetric, that is,  $(x, y) \in D$  implies  $(y, x) \in D$ . For instance,  $D$  might be the set of (binary) questions in a survey.

A population-level dataset is modeled as follows.

**Definition 1.** A *stochastic choice function with response times* (SCF-RT) is a pair of functions  $(p, f)$  where

- (i)  $p$  assigns to each  $(x, y) \in D$  a frequency  $p(x, y) > 0$ , with the property that  $p(x, y) + p(y, x) = 1$ , and
- (ii)  $f$  assigns to each  $(x, y) \in D$  a strictly positive density function  $f(x, y)$  on  $\mathbb{R}_+$ .

In the particular case where  $D$  contains only one pair,  $D = \{(x, y), (y, x)\}$ , we say that  $(p, f)$  is an SCF-RT for  $(x, y)$ .

In our population setting,  $p(x, y)$  is interpreted as the fraction of the population who chose  $x$  when offered the binary choice between  $x$  and  $y$ . The assumption that  $p(x, y) > 0$  for all  $(x, y) \in D$  implies that population choice is non-degenerate, that is, both alternatives enjoy at least some support. The density  $f(x, y)$  describes the distribution of response times conditional on the subset of agents who chose  $x$  in the binary choice between  $x$  and  $y$ . The corresponding cumulative distribution function is denoted by  $F(x, y)$ . The following definition is taken from Alós-Ferrer et al. (2021).

**Definition 2.** A *random utility model with a chronometric function* (RUM-CF) is a triple  $(u, \tilde{v}, r)$  where  $u : X \rightarrow \mathbb{R}$  is a utility function and  $\tilde{v} = (\tilde{v}(x, y))_{(x, y) \in C}$  is a collection of real-valued random variables, with each  $\tilde{v}(x, y)$  having a density function  $g(x, y)$  on  $\mathbb{R}$ , fulfilling the following properties:

$$\text{(RUM.1)} \quad \mathbb{E}[\tilde{v}(x, y)] = u(x) - u(y),$$

$$\text{(RUM.2)} \quad \tilde{v}(x, y) = -\tilde{v}(y, x), \text{ and}$$

(RUM.3) the support of  $\tilde{v}(x, y)$  is connected.

Further,  $r : \mathbb{R}_{++} \rightarrow \mathbb{R}_+$  is a continuous function that is strictly decreasing in  $v$  whenever  $r(v) > 0$ , with  $\lim_{v \rightarrow 0} r(v) = \infty$  and  $\lim_{v \rightarrow \infty} r(v) = 0$ .

The utility function  $u$  is interpreted as a representative utility for a population or group. The random variables  $\tilde{v}(x, y)$  incorporate heterogeneity across individuals, modeled in terms of the distribution of random pairwise utility differences. That is, the density  $g(x, y)$  describes the population distribution of the utility differences between  $x$  and  $y$ , with condition (RUM.1) requiring that the population mean identifies the utilitarian welfare difference. Alternatively,  $g(x, y)$  can be seen as the density of the noise distribution, capturing choice inconsistencies and possible survey bias. In particular, it is important to note that this density might be asymmetric. Condition (RUM.2) reflects that the choice between  $x$  and  $y$  is the same as the choice between  $y$  and  $x$ , and condition (RUM.3) is a regularity condition stating that the distribution of a pair's utility differences has connected support, i.e. without gaps.

Last,  $r$  represents the chronometric function, which maps realized utility differences  $v$  into response times  $r(|v|)$ , such that larger absolute utility differences generate shorter response times, that is, easier choices are faster. This “chronometric effect” is based on extremely well-established empirical regularities from the cognitive sciences (see Alós-Ferrer et al., 2021, for details and references). Given a RUM-CF  $(u, \tilde{v}, r)$  and a pair  $(x, y) \in C$ , the random variable describing the response times predicted by the model conditional on  $x$  being chosen over  $y$  is given by

$$\tilde{t}(x, y) = r(|\tilde{v}(x, y)|),$$

conditional on  $\tilde{v}(x, y) > 0$ .

The results we seek will be in terms of preference revelation for *all* RUM-CFs which rationalize (explain) the data. The following definition pins down the formal meaning of the latter.

**Definition 3.** A RUM-CF  $(u, \tilde{v}, r)$  *rationalizes* an SCF-RT  $(p, f)$  if

- (i)  $p(x, y) = \text{Prob}[\tilde{v}(x, y) > 0]$  holds for all  $(x, y) \in D$ , and
- (ii)  $F(x, y)(t) = \text{Prob}[\tilde{t}(x, y) \leq t \mid \tilde{v}(x, y) > 0]$  holds for all  $t > 0$  and all  $(x, y) \in D$ .

In other words, a RUM-CF (the model) rationalizes an SCF-RT (the data) if it reproduces both the choice frequencies and the conditional response time distributions in the latter. Obviously, fixing the set  $D$ , every RUM-CF generates an SCF-RT through the equations given in (i) and (ii) above, thus an alternative definition is that a RUM-CF

rationalizes an SCF-RT if it coincides with the SCF-RT generated by the former. We say that an SCF-RT is *rationalizable* if there exists a RUM-CF that rationalizes it.<sup>2</sup>

The next definition captures preference revelation in our framework.

**Definition 4.** A rationalizable SCF-RT *reveals a group preference for  $x$  over  $y$*  if all RUM-CFs that rationalize it satisfy  $u(x) \geq u(y)$ . It *reveals a strict group preference for  $x$  over  $y$*  if all RUM-CFs that rationalize it satisfy  $u(x) > u(y)$ .

Our results make use of the following technical concept, taken from Alós-Ferrer et al. (2021). Given two cumulative distribution functions  $G$  and  $H$  on  $\mathbb{R}_+$  and a constant  $q \geq 1$ , we say that  $G$   *$q$ -first-order stochastically dominates  $H$*  (also written  $G$   *$q$ -FSD  $H$* ) if

$$G(t) \leq q \cdot H(t) \text{ for all } t \geq 0.$$

If the inequality is strict for some  $t$ , then  $G$  *strictly  $q$ -first-order stochastically dominates  $H$*  (written  $G$   *$q$ -SFSD  $H$* ). For  $q = 1$ , these concepts coincide with the standard notions of first-order stochastic dominance, but they are weaker when  $q > 1$ . Clearly,  $q$ -FSD implies  $q'$ -FSD whenever  $q \leq q'$ .

The following Theorem is the main result of Alós-Ferrer et al. (2021), reformulated in welfare-utilitarian terms for a population. This result is obviously limited to preference revelation within a single, given group and we state it for completeness and ease of reference.

**Theorem 1.** *A rationalizable SCF-RT  $(p, f)$  reveals a group preference for  $x$  over  $y$  if  $F(y, x)$   $q$ -FSD  $F(x, y)$ , and reveals a strict group preference if  $F(y, x)$   $q$ -SFSD  $F(x, y)$ , for  $q = p(x, y)/p(y, x)$ .*

## 2.2 Group Comparisons

We consider data from two groups,  $A$  and  $B$ , and two fixed choices,  $(x, y)$  and  $(z, w)$ . However, we explicitly allow  $A = B$  (and  $(x, y) \neq (z, w)$ ), in which case the comparison is between two different choices made by the same group. We also explicitly allow  $(x, y) = (z, w)$  when  $A \neq B$ , that is, comparing the same choice across two different groups, and  $z = y$  even when  $A = B$ , that is, comparing two connected choices for the same group, as in staircase designs.

We now adapt all previous concepts to our setting, always taking  $A, B$ , and the pairs  $(x, y)$  and  $(z, w)$  as fixed and given. The data hence must deliver the choice frequencies for  $x$  and  $y$  (for group  $A$ ) and for  $z$  and  $w$  (for group  $B$ ), and the conditional response time distributions.

---

<sup>2</sup>The results in Alós-Ferrer et al. (2021) further study preference revelation within constrained subclasses of RUM-CFs, e.g. the class of all *symmetric* RUM-CFs (meaning that noise terms have symmetric densities). Here we consider only the unrestricted class of all RUM-CFs.

**Definition 5.** A two-group stochastic choice function with response times (2SCF-RT) for choices  $(x, y)$  and  $(z, w)$  is a quadruple  $(p_A, f_A; p_B, f_B)$  such that  $(p_A, f_A)$  is an SCF-RT for the choice  $(x, y)$  and  $(p_B, f_B)$  is an SCF-RT for the choice  $(z, w)$ .

The class of data-generating processes allows for different distributions of utility differences in the two groups, but fixes a chronometric function. Obviously, for applications where  $A = B$ , the latter assumption is void and thus inconsequential. For  $A \neq B$ , this assumption means that  $r$  is taken to reflect a (possibly neurophysiological) relation which is orthogonal to the nature of the groups (although the assumption could be weakened to some extent). We remind the reader, however, that  $r$  is not assumed to have any specific functional form, and that our preference-revelation results below are not only for all utilities and distributions of utilities which rationalize the data, but also for all  $r$ .

**Definition 6.** A two-group random utility model with a chronometric function (2RUM-CF) is a tuple  $(u_A, \tilde{v}_A; u_B, \tilde{v}_B; r)$  such that  $(u_A, \tilde{v}_A, r)$  and  $(u_B, \tilde{v}_B; r)$  are RUM-CFs.

Rationalization is extended in a straightforward way.

**Definition 7.** A 2RUM-CF  $(u_A, \tilde{v}_A; u_B, \tilde{v}_B; r)$  rationalizes a 2SCF-RT  $(p_A, f_A; p_B, f_B)$  for choices  $(x, y)$  and  $(z, w)$  if  $(u_A, \tilde{v}_A, r)$  rationalizes  $(p_A, f_A)$  (for  $(x, y)$ ) and  $(u_B, \tilde{v}_B; r)$  rationalizes  $(p_B, f_B)$  (for  $(z, w)$ ).

The last definition captures our key concept, i.e. when group  $A$  prefers or likes  $x$  over  $y$  more than group  $B$  prefers  $z$  over  $w$ .

**Definition 8.** A rationalizable 2SCF-RT for choices  $(x, y)$  and  $(z, w)$  reveals that group  $A$  prefers  $x$  over  $y$  more than group  $B$  prefers  $z$  over  $w$  if all 2RUM-CFs  $(u_A, \tilde{v}_A; u_B, \tilde{v}_B; r)$  that rationalize it satisfy

$$u_A(x) - u_A(y) \geq u_B(z) - u_B(w).$$

It reveals that group  $A$  prefers  $x$  over  $y$  strictly more than group  $B$  prefers  $z$  over  $w$  if all 2RUM-CFs that rationalize it fulfill the strict version of the inequality above.

An alternative descriptor keeping the utilitarian framework more clearly in sight would be that “the welfare increase from  $y$  to  $x$  for group  $A$  is larger than the welfare increase from  $w$  to  $z$  for group  $B$ ,” but we prefer the nomenclature above for concreteness.

### 2.3 Revealed Cardinal Preference Across Groups or Choice Pairs

Theorem 1 delivers a preference revelation result in our context, which deserves mention because it helps put other results into context. Obviously, if group  $A$  prefers  $x$  over  $y$  but group  $B$  does *not* (strictly) prefer  $z$  over  $w$ , i.e. it rather prefers  $w$  over  $z$ , we can trivially conclude that group  $A$  prefers  $x$  over  $y$  more than group  $B$  prefers  $z$  over  $w$ . This straightforward observation can be translated in terms of revealed preference as follows.

**Corollary 1.** *If a rationalizable 2SCF-RT  $(p_A, f_A; p_B, f_B)$  for choices  $(x, y)$  and  $(z, w)$  satisfies that*

- (a)  $F_A(y, x)$   $q_A$ -FOSD  $F_A(x, y)$ , with  $q_A = p_A(x, y)/p_A(y, x)$ , and
- (b)  $F_B(z, w)$   $(1/q_B)$ -FOSD  $F_B(w, z)$ , with  $q_B = p_B(z, w)/p_B(w, z)$ ,

*then it reveals that group A prefers x over y more than group B prefers z over w.*

To state the following result, define for any SCF-RT  $(p, f)$  including data on a choice  $(a, b)$ ,

$$H(a, b)(t) = p(b, a)F(b, a)(t) - p(a, b)F(a, b)(t)$$

and note that the  $q$ -FOSD property is equivalent to  $H(a, b)(t) \leq 0$  for all  $t$ . Given a rationalizable 2SCF-RT  $(p_A, f_A; p_B, f_B)$ , let  $H_A(x, y)$  and  $H_B(z, w)$  be defined analogously. Then, Corollary 1 states that a sufficient condition for it to be revealed that group A prefers  $x$  over  $y$  more than group B prefers  $z$  over  $w$  is that  $H_A(x, y)(t) \leq 0 \leq H_B(z, w)(t)$  for all  $t$ . The main result in this section identifies a weaker sufficient condition, namely that  $H_A(x, y)(t) \leq H_B(z, w)(t)$  irrespective of whether the sign of either term is constant or not.

**Theorem 2.** *If a rationalizable 2SCF-RT  $(p_A, f_A; p_B, f_B)$  for choices  $(x, y)$  and  $(z, w)$  satisfies that*

$$H_A(x, y)(t) \leq H_B(z, w)(t) \quad \text{for all } t \geq 0,$$

*then it reveals that group A prefers x over y more than group B prefers z over w.*

*If the inequality is strict for some  $t$ , the revealed preference is also strict.*

The proof is in the Appendix. The formalization in terms of  $H_A$  and  $H_B$  makes clear how Theorem 2 generalizes Corollary 1 and hence the implications of Alós-Ferrer et al. (2021). Of course, as the condition becomes more involved, an intuitive interpretation also becomes more difficult. The condition  $H(a, b)(t) \leq 0$  for all  $t$  can be interpreted as stating that revealed errors (choices of  $y$  when, ex post, a preference for  $x$  is revealed) are not much faster in a stochastic dominance sense than revealed correct responses, which very roughly could be taken as a weakening of a “slow errors” condition. Suppose, for concreteness, that  $H_A(x, y)(t) \leq H_B(z, w)(t) \leq 0$ , so that this interpretation holds for group A and group B (and their respective choices) separately. Intuitively,  $H_A(x, y)(t) \leq H_B(z, w)(t)$  for all  $t$  means that the difference is farther away from zero, thus larger, for group A. That is, the condition requires that the difference in response times (actually, the difference in the distributions) is larger for group A than for group B. There are, however, two caveats with this interpretation. The first is that the condition is distributional and cannot be summarized in terms of statistics as, e.g., the mean. The second is that the interpretation is only partial, as it does not take into account that the functions  $H_A, H_B$  incorporate the actual choice frequencies; that is, the actual condition refers to differences between cdf values adjusted by choice frequencies.

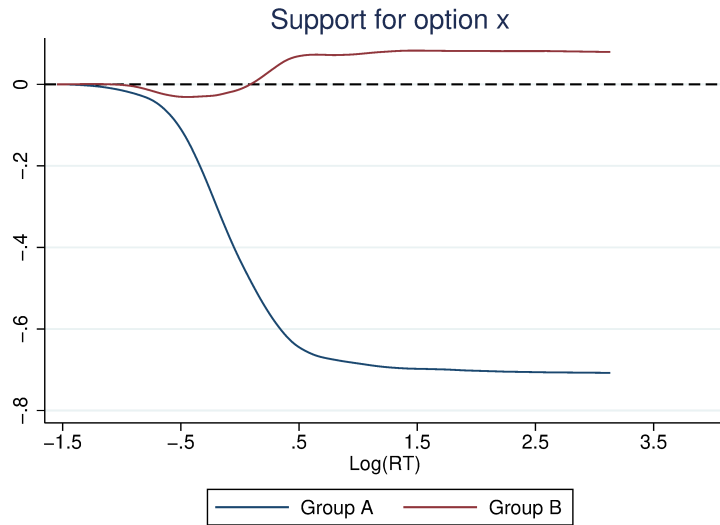


Figure 1: Support for option  $x$  between between group A and group B. If a line does not cross the 0 on the vertical axis, then preferences are revealed according to Theorem 1. If the lines representing the two groups do not cross, then a stronger preference for  $x$  for one group over the other is revealed in agreement with Theorem 2.

## 2.4 An Example

Part of the appeal of the proposed method is how simple it is to apply to a wide array of datasets. One only needs to check the conditions implied by Theorems 1 or 2 in order to reveal preferences. This can be done by means of simple graphical representations.

Theorem 1 intuitively tells us how to weight choice frequencies and response times in order to verify whether preferences are revealed while accounting for noise. In particular, a group is revealed to prefer  $x$  over  $y$  if  $H(x, y)(t) \leq 0$  for all  $t \geq 0$ . To verify this condition the analyst only needs to plot this expression (we detail how to calculate the CDFs below) and check *whether it crosses 0*. If it does, preferences are not revealed, otherwise any model of noise and utility function, within the universe of random utility models, will deliver the same preference ordering. For example, in Figure 1, preferences for option  $x$  are revealed for Group A, but not for Group B.

Theorem 2 tells us under which conditions one can say that a group has a stronger preference for one option over another than a different group. In particular, Group A is revealed to prefer  $x$  over  $y$  more than Group B if  $H_A(x, y)(t) \leq H_B(x, y)(t)$  for all  $t \geq 0$ . To verify this condition the analyst only needs to plot the same expressions required for Theorem 1 and check *whether they intersect*. If they do, preferences across groups are not revealed, otherwise any model of noise and utility function, within the universe of random utility models, will deliver the ordering of utility differences. For example, in Figure 1, Group A has a stronger preference for option  $x$  than Group B.

## 3 Application to a Representative Survey

### 3.1 Survey Design

We ran a pre-registered, online survey with a representative sample of the UK population ( $N = 1202$ ; see Appendix B for the derivation of the required sample size).<sup>3</sup> The representativeness of the sample is in terms of gender, age, and ethnicity. The survey, conducted through Prolific, lasted around five minutes and participants were compensated with a flat fee of £0.5. Participants provided binary answers to 23 questions on a wide range of socio-economic topics, e.g. support for gender quotas, redistributive policies, mandatory vaccinations, and greenhouse gas policies. The topics of the implemented questions are not central to the scope of the article, but their wide range and the relevance of the issues is used to showcase the applicability and implications of the proposed method.

To apply our method, for each question we divide participants into two groups according to a dimension relevant to the question, e.g. gender, age, political orientation (left vs. right), or income. For comparison purposes, we then investigate differences in the proportion of people supporting one of the two options between the groups by means of a simple test of proportions (that only uses choice data). We then apply Theorem 2 (hence using both choices and response times) to examine whether a stronger preference for an option in one group compared to the other group is revealed, and in particular whether preference revelation across groups can be obtained even though the corresponding test of proportions is not significant. Conversely, we also identify cases where tests of proportions are significant but preference revelation across groups is unwarranted.

We also apply Theorem 1 to examine when preferences for one option against another are revealed for each group, and when they are not. This allows us to show that even in cases where preferences are not revealed for one of the groups, Theorem 2 can deliver a ranking of preferences between the groups. That is, there are situations where one cannot conclude that a group prefers an option over the other, but we can still say that a group prefers an option more than the other group.

For each survey question, we choose a dimension to define the groups that is likely to lead to polarized opinions (the dimension to be applied to each question was preregistered). This stacks the odds in favor of simple tests of proportions finding significant differences, and hence against our method providing new evidence. The complete list of questions and chosen group divisions is reported in Appendix C.

### 3.2 Survey Analysis

In the following subsections, we showcase the most interesting empirical results from the survey. Those are selected to illustrate the different possibilities that can arise in

---

<sup>3</sup>Preregistration: AEA RCT Registry AEARCTR-0009022. Ethical approval: OEC IRB 2022-012 (The Human Subjects Committee of the Faculty of Economics, Business Administration and Information Technology at the University of Zurich).



the application of the method. The entire list of results for our survey is presented in Appendix D. The technical details of the estimation of response time distributions are as in Alós-Ferrer et al. (2021) for the empirical revelation of preferences.<sup>4</sup> The key difference is that Alós-Ferrer et al. (2021) performed estimations at the individual level, requiring multiple answers for the same question for each individual, while we conduct estimations at the group level, which only requires a single choice (and response time) from each participant.

First, we observe that the condition in Theorem 1 is fulfilled quite often, that is, group preferences are revealed very frequently. Of the 46 possibilities we examine (23 questions for two groups each), Theorem 1 reveals preferences in 38 cases (82.61%). This is a first validation of the interest of the techniques.

Concerning revelation of preference differences across groups (Theorem 2), we observe the following cases. For 15 of the 23 questions (65.22%), the conditions of Theorem 2 are fulfilled (the  $H$  functions do not cross), and hence we obtain preference revelation across groups. For 10 of those questions, tests of proportions are also statistically significant. In these cases, our method agrees with choice-based tests, but provides a stronger result, since preference revelation through Theorem 2 is independent of any assumptions on the distribution of noise of the shape of underlying utilities.

In five cases (overall 21.74%), preference revelation across groups obtains even though tests of proportions yield no significant differences. In these cases, our method (that uses more data than tests based on choice frequencies only) provides new information. There are two possible reasons why these cases might occur. It might be that the preferences of two different groups are systematically different but relatively close. Hence, when choices are noisy it is likely that statistical tests accept the null hypothesis, delivering a *false negative* result. By using an additional source of information (response times), the analyst effectively reduces measurement errors and is able to identify a difference between the two groups. However, for  $N = 1,202$  and a significance level of  $\alpha = 0.05$ , with a small effect size (Cohen's  $d \leq 0.2$ ), the expected proportion of false negatives is below 0.1%, thus this explanation is unlikely to apply to all five cases. A more likely explanation is that the preferences of the two groups are different, but many individuals of one group are affected by social desirability bias and misrepresent their preferences in the survey.

For the remaining eight questions (34.78%), the  $H$  functions cross, hence there is no preference revelation across groups. In four of those cases, tests of proportions also yield no significant differences. In the remaining four cases (overall 17.39%), tests of proportions are significant even though preference revelation does not obtain. These cases should be interpreted with care. Since Theorem 2 identifies a *sufficient condition* for preference revelation, it might simply be that preferences are indeed different across

---

<sup>4</sup>In particular, we estimate the distribution of log-transformed response times to avoid boundary problems. The estimates use an Epanechnikov kernel with optimally chosen non-adaptive bandwidth. The kernel density estimates were performed in *Stata* using the *akdensity* function, which delivers CDFs as output.

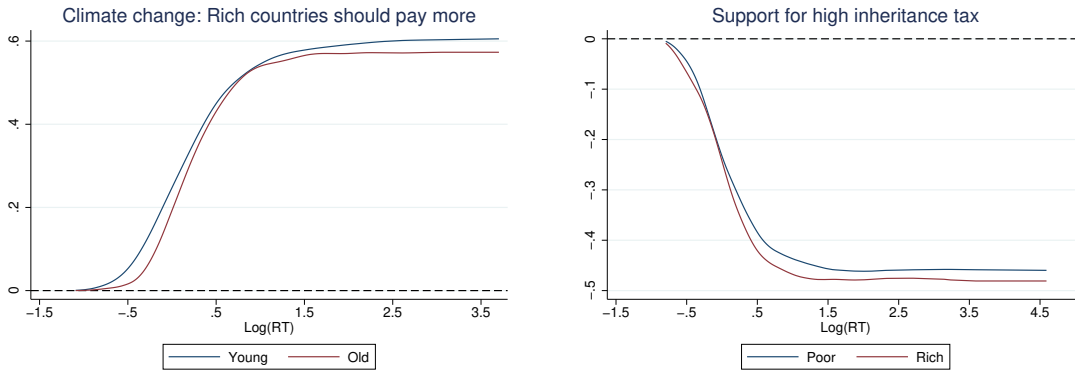


Figure 2: Support for policies which allocate higher costs to rich countries compared to poor ones between younger vs. older participants (left) and support for a high inheritance tax between income range (right) represented by the  $H$  functions plotted against log-transformed response times.

groups, but the condition is not fulfilled. On the other hand, it might be that group preferences are actually not different, and the test of proportions simply reflects survey bias.

Obviously, further comparisons can be carried out exploring different combinations of questions and groups. However, the purpose of the survey was not to investigate how often statistical significance is unwarranted, or how often a revealed difference in preferences is obtained when statistical tests fail. Rather, our objective was to illustrate that, in the presence of survey bias, preference revelation across groups is a different question than the comparison of choice proportions, and can be achieved by means of inexpensive, easy-to-apply techniques.

### 3.3 Getting More from the Data

The left-hand panel of Figure 2 depicts the support for the proposition that rich countries should pay more than poor countries to finance policies aimed at decreasing greenhouse gas emissions (independently of the individual country history of emissions). Comparing the support for this proposition between younger vs. older participants shows no statistically significant differences (Young: 80.40% vs. Old: 78.67%, test of proportions,  $N = 1202$ ,  $z = 0.7442$ ,  $p = 0.457$ ). However, the conditions of Theorem 2 are fulfilled (the two  $H$  functions do not intersect), as are the conditions of Theorem 1 for each group (neither  $H$  function crosses the zero). This means that both groups support this policy, but young people do support it more than older people. This result suggests that the lack of a statistical significance is either a false negative result or the consequence of survey bias causing older people to state (their perception of) a more socially-desirable response. Our method flags hence allows the researcher to arrive at a positive conclusion even though the statistical test is inconclusive.

The right-hand panel of Figure 2 depicts another example of this phenomenon. In particular, it represents the support for a high inheritance tax between a median split of income. This is another case where comparing response proportions between poor vs. rich people suggests no statistically significant differences (Poor: 26.98% vs. Rich: 25.96%, test of proportions,  $N = 1202$ ,  $z = 0.3675$ ,  $p = 0.713$ ). However, none of the  $H$  functions crosses the zero, meaning preferences are revealed for both groups (Theorem 1). Also, the  $H$  functions do not cross, meaning that poorer participants have stronger preferences for high inheritance taxes compared to richer participants.

Among our 23 questions we find that five comparisons fall within this category. That is, we find 21.74% of cases where a researcher might be misled to conclude that there is no difference by a non-significant test result, while our method reveals a preference difference. The other three examples are as follows (see Appendix D.1). First, both genders reveal a preference against women earning less than men for the same job, but the preference of females is revealed to be stronger than that of males. Second, poorer people exhibit a stronger preference for sharply increasing taxation for the upper 1% than richer people (and both groups reveal a preference in favor). Third, both the poorer and the richer reveal a preference against a hypothetical policy increasing their personal yearly income by £2500 at the cost of a reduction of £5000 in the average yearly income in the U.K., but the richer exhibit a stronger preference in favor than the poorer.

### 3.4 Unwarranted Conclusions

The left-hand panel of Figure 3 depicts the support for the proposition that policies aimed at reducing greenhouse gas emissions should be financed by rich countries only, instead of by all countries proportionally to current emissions. For this question, comparing the response proportions between age groups suggests a statistically significant result (Young: 27.91% vs. Old: 23.50%, test of proportions,  $N = 1202$ ,  $z = 1.748$ ,  $p = 0.040$ ). However, while Theorem 1 reveals preferences for both groups (i.e., the  $H$  functions do not cross zero), the two  $H$  functions intersect, and hence Theorem 2 does not reveal group preference differences. Thus, a researcher would not be entitled to conclude that younger people support this policy more than older people. This result suggests that the statistical significance might be an artifact of survey bias (e.g., some younger people stating a preference more aligned with a perception of social desirability), which our method flags, preventing the researcher from arriving at unwarranted conclusions.

The right-hand panel of Figure 3 depicts another example of this phenomenon. It represents the support for gender quotas between genders. This is another case where comparing response proportions between genders suggests a (highly) statistically significant result (Females: 34.15% vs. Males: 24.70%, test of proportions,  $N = 1202$ ,  $z = 3.588$ ,  $p < 0.001$ ). However, the two  $H$  functions intersect, hence revealing that a researchers would actually not be warranted to conclude that women are more in favor of gender quotas than man.

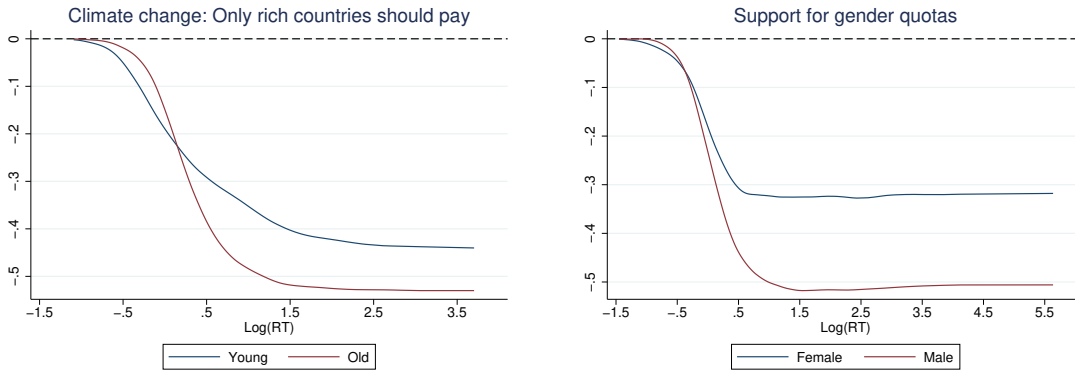


Figure 3: Support for the proposition that policies aimed at reducing greenhouse gas emissions should be financed by rich countries only, between age groups (left), and support for gender quotas, between genders (right), represented by the  $H$  functions plotted against log-transformed response times.

Among our 23 questions we find that four comparisons fall within this category. That is, in 17.39% of cases a researcher might incorrectly conclude that there is a difference in preferences between the groups, when the data does not actually substantiate the claim. The other two examples are as follows (see Appendix D.2). First, a test of proportions shows larger support among the young (compared to the old) for the proposition that greenhouse-reducing policies should impose larger costs in the future, but again Theorem 2 detects no preference differences (and both groups oppose it according to Theorem 1). Second, a test of proportions shows larger support for a hypothetical policy decreasing the respondent’s personal yearly income by £2500 in exchange for an increase in the average yearly income of £5000 in the U.K., but Theorem 2 detects no preference differences (and both groups oppose it according to Theorem 1).

### 3.5 Positive Concordance

The left-hand panel of Figure 4 depicts the support for accepting more immigrants in the U.K. For this question, comparing response proportions between political orientations detects a statistically significant difference (Left: 74.06% vs. Right: 33.41%, test of proportions,  $N = 1202$ ,  $z = 13.488$ ,  $p < 0.001$ ). Theorems 1 and 2 show that preferences for both groups are revealed ( $H$  functions do not cross zero), and, naturally, that people on the left of the political spectrum have a stronger preference for supporting additional immigration than those on the right. Actually, since preferences for the groups are revealed in opposite directions, this is a case where the difference between the groups follows from Corollary 1.

The left-hand panel of Figure 4 depicts the willingness to cheat on taxes (“Would you cheat on taxes if you had a chance?”). Comparing response proportions across genders again detects a statistically significant difference (Male: 30.02% vs. Female: 21.43%, test of proportions,  $N = 1202$ ,  $z = 3.315$ ,  $p < 0.001$ ). Theorem 1 shows that both groups

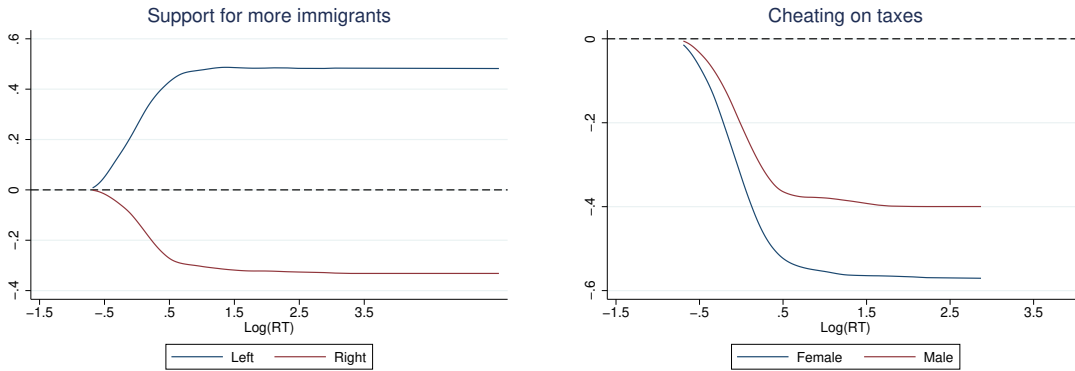


Figure 4: Support for immigration between political orientations (left) and willingness to cheat on taxes between genders (right), represented by the  $H$  functions plotted against log-transformed response times.

express a preference against cheating on taxes ( $H$  functions do not cross zero). However, 2 confirms the gender difference: males have a weaker preference against cheating on taxes than females.

Among our 23 questions, we find that 10 comparisons fall within this category (see Appendix D.3). For example, respondents on the right think that people get ahead mostly because of their own merits, while people on the left think it is mostly due to luck and the help of others (this is another example of Corollary 1). In other examples, both groups reveal the same preference according to Theorem 1 and Theorem 2 additionally reveals a difference in preferences. For example, both people on the left and on the right oppose limiting international trade to protect national jobs, but this opposition is weaker for people on the left. Both genders express a preference against cheating on a partner, but this preference is revealed to be weaker for men. Both younger and older respondents reveal a preference in favor of taxes on emissions, but this preference is stronger for the young. Both respondents on the left and on the right favor guaranteed access to health services for people who do not work, but this preference is stronger for respondents on the left.

### 3.6 Negative Concordance

The left-hand panel of Figure 5 depicts the support for a mandatory COVID vaccine. For this question, comparing response proportions between political orientations leads to accept the null hypothesis (Left: 43.80% vs. Right: 43.76%, test of proportions,  $N = 1202$ ,  $z = 0.0119$ ,  $p = 0.991$ ). Theorem 1 shows that, for both groups, preferences against mandatory vaccines are revealed ( $H$  functions do not cross zero). However, the two  $H$  functions intersect, and hence Theorem 2 does not reveal a difference in preferences across groups. It is hence unwarranted to conclude that people on the left of the political spectrum support this policy more than those on the right.

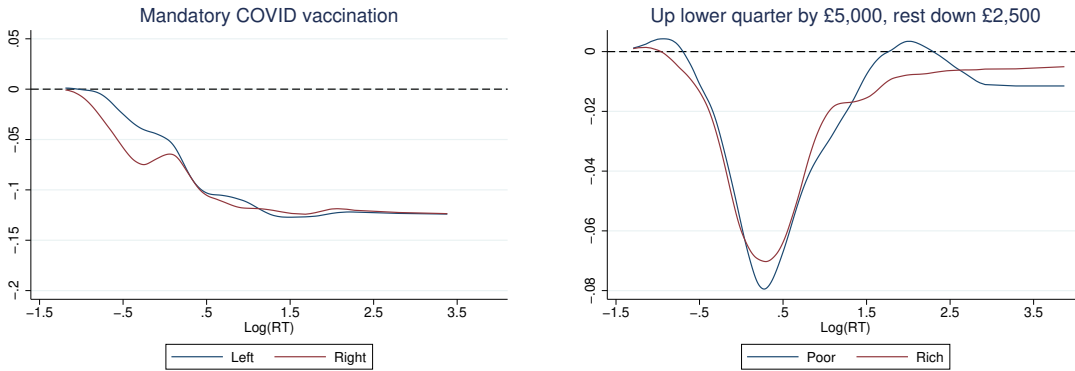


Figure 5: Support for immigration (left) and for a mandatory COVID vaccine (right) between political orientation represented by the  $H$  functions plotted against log-transformed response times.

The right-hand panel of Figure 5 depicts the support for a hypothetical policy that would increase the yearly income of the lower quarter of the income distribution by £5000, while decreasing the yearly income of the rest by £2,500. For this question, comparing response proportions between political orientations again leads to accept the null hypothesis (Rich: 49.78% vs. Poor: 49.43%, test of proportions,  $N = 1202$ ,  $z = 0.085$ ,  $p = 0.933$ ). In this case, preference is not revealed for either group according to Theorem 1, and a difference in preferences is also not revealed by Theorem 2.

Among our 23 questions we find that four comparisons fall within this category. The other two questions (see Appendix D.4) are as follows. First, both males and females express a preference against mothers staying at home, but there is no revealed preference difference. Second, both poorer and richer respondents reveal supporting the proposition that reducing inequality should be the government’s responsibility, but a difference across groups is also not revealed.

### 3.7 Revealed Group Differences when Preferences Are not Revealed

In this section we illustrate that even in cases where preferences are not revealed for one of the groups according to Theorem 1 it can be still the case that Theorem 2 delivers a ranking of preferences between the groups.

For example, as illustrated in Figure 6 when examining support for Brexit (“Do you think Brexit was a good idea?”), Theorem 1 reveals a preference against Brexit for people on the left of the political spectrum, but no preference is revealed for those on the right. However, when comparing groups, using Theorem 2 we can still conclude that those on the right support Brexit more than those on the left. This is actually a case of positive concordance, as the test of proportions is significant (Right: 53.97% vs. Left: 14.59%, test of proportions,  $N = 1202$ ,  $z = 14.497$ ,  $p < 0.001$ ).

The same phenomenon occurred for weed legalization, where younger people reveal a preference in favor, the preference of older people is not revealed, but a difference across

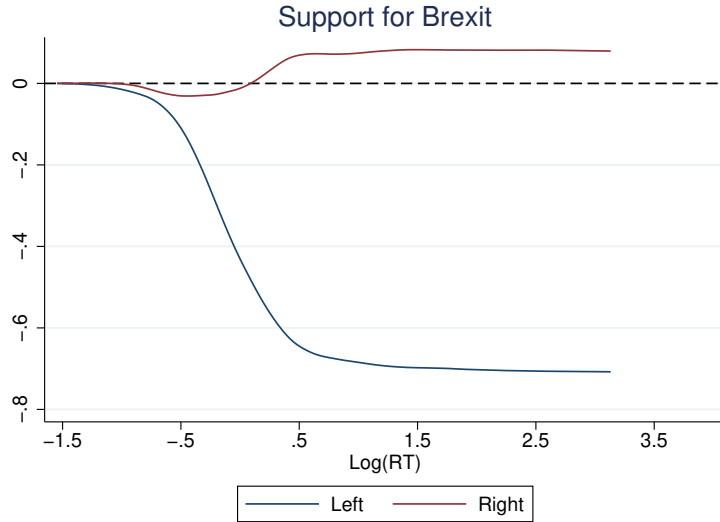


Figure 6: Support for Brexit between people reporting being more on the left compared to the right of the political spectrum represented by the  $H$  functions plotted against log-transformed response times.

groups is revealed. Analogously, support for same-sex marriage is revealed for the left but not for the right, and a difference across groups is revealed. These are two further examples of positive concordance (Appendix D.3). A further example concerns support for a hypothetical redistribution increasing the own yearly personal income by £2500 while reducing the average by £5000, where the poor reveal a preference against, no preference is revealed for the rich, and a difference across the groups is revealed. The latter is a case where the corresponding test of proportions is not significant (Appendix D.1).

## 4 Uniformly Revealed Preferences

In this section, we consider a stronger concept of preference revelation across groups. By strengthening the concept, we obtain a stronger analytical result, namely a full characterization in the form of a necessary and sufficient condition on the dataset. The price to pay is that this condition is far more demanding than the sufficient one in Theorem 2, and as a consequence it is fulfilled empirically far less often.

### 4.1 Analytical Result

The following definition spells out the stronger concept.

**Definition 9.** A rationalizable 2SCF-RT reveals that group  $A$  prefers  $x$  over  $y$  uniformly more than group  $B$  prefers  $z$  over  $w$  if, for each 2RUM-CF  $(u_A, \tilde{v}_A; u_B, \tilde{v}_B; r)$  that rationalizes it,  $G_A(x, y)$  FOSD  $G_B(z, w)$ , where  $G_A(x, y)$  and  $G_B(z, w)$  are the cumulative distribution functions of  $\tilde{v}_A$  and  $\tilde{v}_B$ , respectively.

If a preference is uniformly revealed in this sense, the preference of group  $A$  for  $x$  over  $y$ , in terms of cardinal utility differences  $\tilde{v}_A(x, y)$  for group  $A$  are systematically larger (in the stochastic dominance sense) than the utility differences  $\tilde{v}_B(z, w)$ . Of course, this implies that the means are ordered in the same way,  $u_A(x) - u_A(y) \geq u_B(z) - u_B(w)$ , and thus uniformly revealed preference (Definition 9) implies revealed preference (Definition 8).

The following result shows that a sufficient condition ensuring uniform preference revelation is also sufficient, that is, strengthening the preference revelation concept allows us to obtain a full characterization.

**Theorem 3.** *A rationalizable 2SCF-RT  $(p_A, f_A; p_B, f_B)$  reveals that group  $A$  prefers  $x$  over  $y$  uniformly more than group  $B$  prefers  $z$  over  $w$  if and only if the following two conditions hold.*

- (a)  $F_A(y, x) \text{ } p_B(w, z) / p_A(y, x)$ -FOSD  $F_B(w, z)$ , and
- (b)  $F_B(z, w) \text{ } p_A(x, y) / p_B(z, w)$ -FOSD  $F_A(x, y)$ .

Further, it is easy to show that if the true data generating process does fulfill that  $G_A(x, y)$  FOSD  $G_B(z, w)$ , then the data *must* fulfill the conditions in Theorem 3.

**Corollary 2.** *Consider a 2RUM-CF such that  $G_A(x, y)$  FOSD  $G_B(z, w)$ . Then, the generated 2SCF-RT fulfills (a) and (b) in Theorem 3 and hence reveals that group  $A$  prefers  $x$  over  $y$  uniformly more than group  $B$  prefers  $z$  over  $w$ .*

The interpretation of the conditions in Theorem 3 is as follows. Ex post, it is revealed that group  $A$  prefers  $x$  over  $y$  (uniformly) more than group  $B$  prefers  $z$  over  $w$ . Although this does not imply that a preference for  $x$  over  $y$  is revealed for group  $A$ , nor that a preference for  $z$  over  $w$  is revealed for group  $B$ , suppose for concreteness that this is the case. Hence, choices of  $y$  or  $w$  are revealed errors and choices of  $x$  or  $z$  are revealed correct answers. Thus, (a) states that revealed errors of group  $A$  are not much faster than revealed errors of group  $B$ , and condition (b) states that revealed correct answers of group  $A$  are not much faster than revealed correct answers of group  $B$ .

The meaning of “not much faster” is measured in terms of choice frequencies in the following sense. Suppose that choice frequencies (for  $x$  and  $z$ , respectively) are identical in group  $A$  and group  $B$ . Then, the information coming from choice frequencies does not help at all to differentiate the preferences of the groups, and conditions (a) and (b) become regular FOSD statements, i.e. “not much faster” means “slower in the FOSD sense.” That is, in the absence of useful information from choice frequencies, the conditions on response times which ensure (uniform) preference revelation become strong. On the opposite extreme, if the choice frequency of  $x$  relative to the choice frequency of  $z$  becomes very large, conditions (a) and (b) become weaker, and in the limit are always vacuously fulfilled. That is, if choice frequencies are very informative, very little information from response times is needed to obtain (uniform) preference revelation.



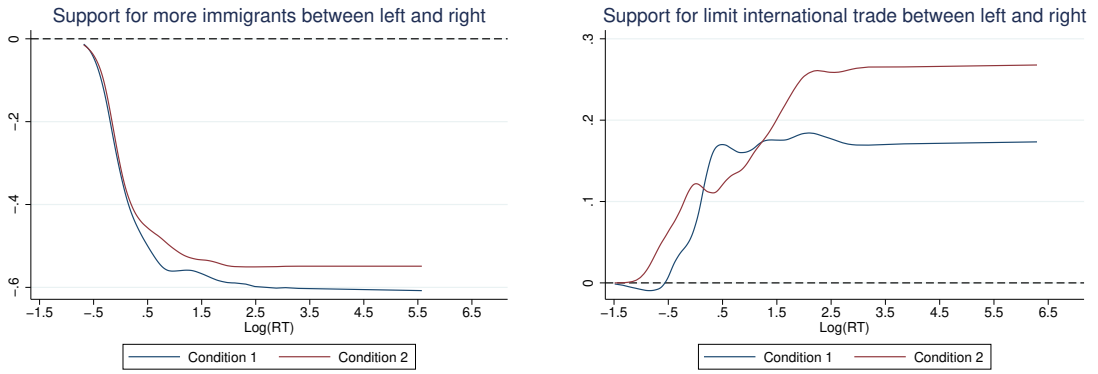


Figure 7: Support for allowing more immigrants (left) and for limiting international trade (right) between political orientation represented by the two conditions of Theorem 3.

## 4.2 Empirical Analysis

We apply the more stringent Theorem 3 to our survey data. Since uniform preference revelation implies preference revelation, it follows that the condition of Theorem 3 must fail whenever Theorem 2 failed. Hence, we only need to consider the set of 15 questions where Theorem 2 revealed a preference difference across groups.

Of those questions, however, only for two (8.70% of the questions in the survey) uniform preference differences are revealed according to Theorem 3. Figure 7 gives a graphical representation of a case where preferences are revealed according to Theorem 3 (left) and one where they are not (right). The plotted lines are the functions

$$p_A(y, x)F_A(y, x) - p_B(w, z)F_B(w, z) \quad \text{and} \quad p_B(z, w)F_B(z, w) - p_A(x, y)F_A(x, y)$$

and hence differ from all earlier graphical illustrations. Uniform preference revelation according to Theorem 3 occurs if and only if both functions are either always positive or always negative.

As shown in Figure 7 (left), people on the left of the political spectrum (uniformly) support additional immigration more than those on the right.<sup>5</sup> On the contrary, as shown on the right-hand side of the figure, there is no uniform preference revelation between the left and the right for limiting international trade, even though a preference difference was revealed by Theorem 2.

Given the small proportion of cases where the conditions in Theorem 3 bite, and given that these conditions are both necessary and sufficient, the empirical analysis suggests that the concept of uniform preference revelation might be too stringent.

<sup>5</sup>The only other question where Theorem 3 identifies a difference is whether people get ahead because of their own merits.

## 5 Conclusion

Surveys are an essential instrument to elicit societal preferences in a large variety of economic, political, and social issues, and are also regularly used to uncover differences in preferences across different socioeconomic or demographic groups. However, survey data is notoriously noisy, and survey bias is ubiquitous, strongly limiting the reliability and usefulness of standard analyses.

In this work, we have presented a new way to analyze survey data to actually *reveal* preferences of groups and preference differences across groups. We do so by relying on response times, which are both inexpensive and easily-collected in the digital age. Our results are obtained by incorporating insights from psychology and neuroscience in standard economic models of noisy choice. We provide ready-to-use techniques which can uncover group preferences and preference differences even when standard statistical tests are inconclusive. The reason, in addition to the fact that our techniques use more data than those tests, is that the question we answer is a different one. In the presence of survey bias, the question that the analyst should answer is not whether a majority of people state that they support a certain proposition, but rather whether the data (and *all* the dimensions of data) allow to reveal an actual preference (or preference difference) between the alternatives.

In addition to deriving theoretical conditions for preference revelation, in this work we have validated the approach and illustrated its usefulness in a representative, pre-registered survey using a large variety of relevant questions, ranging from support for vaccine mandates to redistributive policies, and from the financing of policies reducing greenhouse-gas emissions to whether one would cheat on taxes or on a partner. The results confirm that the new techniques are useful and often deliver new insights.

The techniques developed in this work are ready for immediate application and have the potential to change how survey data is analyzed in economics, political science, marketing, health research, and many other fields.

## References

- Alesina, A. and N. Fuchs-Schündeln (2007). Good-Bye Lenin (or Not?): The Effect of Communism on People’s Preferences. *American Economic Review* 97(4), 1507–1528.
- Alós-Ferrer, C., E. Fehr, and M. Garagnani (2022). Identifying Nontransitive Preferences. Discussion Paper.
- Alós-Ferrer, C., E. Fehr, and N. Netzer (2021). Time Will Tell: Recovering Preferences when Choices are Noisy. *Journal of Political Economy* 129(6), 1828–1877.
- Alós-Ferrer, C. and M. Garagnani (2022a). Strength of Preference and Decisions Under Risk. *Journal of Risk and Uncertainty* 64(3), 309–329.
- Alós-Ferrer, C. and M. Garagnani (2022b). The Gradual Nature of Economic Errors. *Journal of Economic Behavior and Organization* 200, 55–66.

- Anderson, S. P., J.-F. Thisse, and A. De Palma (1992). *Discrete Choice Theory of Product Differentiation*. Cambridge, MA: MIT Press.
- Baldassi, C., S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci (2020). A Behavioral Characterization of the Drift Diffusion Model and its Multi-Alternative Extension to Choice under Time Pressure. *Management Science* 66(11), 5075–5093.
- Baltas, G. and P. Doyle (2001). Random Utility Models in Marketing Research: A Survey. *Journal of Business Research* 51(2), 115–125.
- Bechtel, M. M., F. Genovese, and K. F. Scheve (2019). Interests, Norms and Support for the Provision of Global Public Goods: The Case of Climate Cooperation. *British Journal of Political Science* 49(4), 1333–1355.
- Bechtel, M. M. and R. Liesch (2020). Reforms and Redistribution: Disentangling the Egoistic and Sociotropic Origins of Voter Preferences. *Public Opinion Quarterly* 84(1), 1–23.
- Bechtel, M. M. and K. F. Scheve (2013). Mass Support for Global Climate Agreements Depends on Institutional Design. *Proceedings of the National Academy of Sciences* 110(34), 13763–13768.
- Bechtel, M. M., K. F. Scheve, and E. van Lieshout (2020). Constant Carbon Pricing Increases Support for Climate Action Compared to Ramping up Costs Over Time. *Nature Climate Change* 10(11), 1004–1009.
- Bertrand, M. and S. Mullainathan (2001). Do People Mean What They Say? Implications for Subjective Survey Data. *American Economic Review (Papers & Proceedings)* 91(2), 67–72.
- Block, H. D. and J. Marschak (1960). Random Orderings and Stochastic Theories of Responses. In I. Olkin (Ed.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, pp. 97–132. Stanford: Stanford University Press.
- Bursztyn, L., A. L. González, and D. Yanagizawa-Drott (2020). Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia. *American Economic Review* 110(10), 2997–3029.
- Camerer, C. F. (1989). Does the Basketball Market Believe in the ‘Hot Hand’. *American Economic Review* 79, 1257–1261.
- Dashiell, J. F. (1937). Affective Value-Distances as a Determinant of Aesthetic Judgment-Times. *American Journal of Psychology* 50, 57–67.
- d’Aspremont, C. and L. Gevers (2002). Social Welfare Functionals and Interpersonal Comparability. In K. J. Arrow, A. K. Sen, and K. Suzumura (Eds.), *Handbook of Social Choice and Welfare*, pp. 450–541. Elsevier.
- Day, R. L. (1968). Preference Tests and the Management of Product Features. *Journal of Marketing* 32(3), 24–29.
- de Quidt, J., J. Haushofer, and C. Roth (2018). Measuring and Bounding Experimenters Demand. *American Economic Review* 108(11), 3266–3302.

- Enke, B., R. Rodríguez-Padilla, and F. Zimmermann (2022). Moral Universalism and the Structure of Ideology. *Review of Economic Studies* forthcoming.
- Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018). Global Evidence on Economic Preferences. *Quarterly Journal of Economics* 133(4), 1645–1692.
- Fehr, E., T. Epper, and J. Senn (2022). Other-Regarding Preferences and Redistributive Politics. Working Paper.
- Feng, Y., R. Caldentey, and C. T. Ryan (2022). Robust Learning of Consumer Preferences. *Operations Research* 70(2), 918–962.
- Fisman, R., S. S. Iyengar, E. Kamenica, and I. Simonson (2006). Gender Differences in Mate Selection: Evidence from a Speed Dating Experiment. *The Quarterly Journal of Economics* 121(2), 673–697.
- Fisman, R., S. S. Iyengar, E. Kamenica, and I. Simonson (2008). Racial Preferences in Dating. *The Review of Economic Studies* 75(1), 117–132.
- Frank, R. H. (2016). *Success and Luck*. Princeton: Princeton University Press.
- Fudenberg, D., P. Strack, and T. Strzalecki (2018). Speed, Accuracy, and the Optimal Timing of Choices. *American Economic Review* 108(12), 3651–3684.
- Funk, P. (2016). How Accurate Are Surveyed Preferences for Public Policies? Evidence From a Unique Institutional Setup. *Review of Economics and Statistics* 98(3), 442–454.
- Giglio, S., M. Maggiori, J. Stroebel, and S. Utkus (2021). Five Facts about Beliefs and Portfolios. *American Economic Review* 111(5), 1481–1522.
- Gillen, B., E. Snowberg, and L. Yariv (2019). Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study. *Journal of Political Economy* 127(4), 1826–1863.
- Goddard, M. and P. Smith (2001). Equity of Access to Health Care Services: Theory and Evidence from the UK. *Social Science & Medicine* 53(9), 1149–1162.
- Hainmueller, J., D. Hangartner, and T. Yamamoto (2015). Validating Vignette and Conjoint Survey Experiments Against Real-World Behavior. *Proceedings of the National Academy of Sciences* 112(8), 2395–2400.
- Hey, J. D. and C. Orme (1994). Investigating Generalizations of Expected Utility Theory Using Experimental Data. *Econometrica* 62(6), 1291–1326.
- Karp, J. A. (2009). Candidate Effects and Spill-over in Mixed Systems: Evidence from New Zealand. *Electoral Studies* 28(1), 41–50.
- Laming, D. (1985). Some Principles of Sensory Analysis. *Psychological Review* 92(4), 462–485.
- Leeper, T. J., S. B. Hobolt, and J. Tilley (2020). Measuring Subgroup Preferences in Conjoint Experiments. *Political Analysis* 28(2), 207–221.

- Lo, A. (2019). Demystifying the Integrated Tail Probability Expectation Formula. *The American Statistician* 73(4), 367–374.
- Loomes, G., P. G. Moffatt, and R. Sugden (2002). A Microeconomic Test of Alternative Stochastic Theories of Risky Choice. *Journal of Risk and Uncertainty* 24(2), 103–130.
- Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.
- Luce, R. D. and J. W. Tukey (1964). Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement. *Journal of Mathematical Psychology* 1(1), 1–27.
- McFadden, D. L. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. New York: Academic Press.
- McFadden, D. L. (2001). Economic Choices. *American Economic Review* 91(3), 351–378.
- Moyer, R. S. and T. K. Landauer (1967). Time Required for Judgements of Numerical Inequality. *Nature* 215(5109), 1519–1520.
- Nownes, A. J. (1992). Primaries, General Elections, and Voter Turnout: A Multinomial Logit Model of the Decision to Vote. *American Politics Quarterly* 20(2), 205–226.
- Ortoleva, P., E. Safonov, and L. Yariv (2021). Who Cares More? Allocation with Diverse Preference Intensities. National Bureau of Economic Research.
- Shadlen, M. N. and R. Kiani (2013). Decision Making as a Window on Cognition. *Neuron* 80, 791–806.
- Sides, J. and J. Citrin (2007). European Opinion About Immigration: The Role of Identities, Interests and Information. *British Journal of Political Science* 37(3), 477–504.
- Snowberg, E. and L. Yariv (2021). Testing the Waters: Behavior Across Participant Pools. *American Economic Review* 111(2), 687–719.
- Thurstone, L. L. (1927). A Law of Comparative Judgement. *Psychological Review* 34, 273–286.
- Tversky, A. (1969). Intransitivity of Preferences. *Psychological Review* 76, 31–48.
- Vakratsas, D. and T. Ambler (1999). How Advertising Works: What Do We Really Know? *Journal of Marketing* 63(1), 26–43.
- Zaller, J. and S. Feldman (1992). A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences. *American Journal of Political Science* 36(3), 579–616.
- Zizzo, D. J. (2010). Experimenter Demand Effects in Economic Experiments. *Experimental Economics* 13, 75–98.

## Appendix A Proofs

*Proof of Corollary 1.* By Theorem 1, for any 2RUM-CF  $(u_A, \tilde{v}_A; u_B, \tilde{v}_B; r)$  that rationalizes  $(p_A, f_A; p_B, f_B)$ ,

$$u_A(x) \geq u_A(y) \quad \text{and} \quad u_B(w) \geq u_B(z)$$

and thus  $u_A(x) - u_A(y) \geq 0 \geq u_B(z) - u_B(w)$ .  $\square$

The following lemmata will be useful below.

**Lemma 1.** *If an SCF-RT  $(p, f)$  including data on a choice  $(a, b)$  is rationalized by a RUM-CF  $(u, \tilde{v}, r)$ , then*

$$H(a, b)(t) = G(a, b)(r^{-1}(t)) + G(a, b)(-r^{-1}(t)) - 1.$$

*Proof of Lemma 1.* Let  $g(a, b)$  be the density function of  $\tilde{v}(a, b)$  and  $G(a, b)$  denote the corresponding cumulative distribution function. By Definitions 1, 2, and 3,  $p(b, a) = G(a, b)(0)$ ,  $p(a, b) = 1 - G(a, b)(0)$ ,  $F(a, b) = (1 - G(a, b)(r^{-1}(t)))/(1 - G(0))$ , and  $F(b, a)(t) = G(a, b)(-r^{-1}(t))/G(a, b)(0)$ . Thus,

$$H(a, b)(t) = G(a, b)(-r^{-1}(t)) - (1 - G(a, b)(r^{-1}(t))) = G(a, b)(r^{-1}(t)) + G(a, b)(-r^{-1}(t)) - 1.$$

$\square$

**Lemma 2.** *For any 2RUM-CF  $(u_A, \tilde{v}_A; u_B, \tilde{v}_B; r)$ ,*

$$E[\tilde{v}_A(x, y)] - E[\tilde{v}_B(z, w)] = \int_{-\infty}^{+\infty} [G_B(z, w)(v) - G_A(x, y)(v)] dv.$$

*Proof of Lemma 2.* By the integrated tail formula for expectations (Lo, 2019), if  $F$  is the cumulative distribution function of the real-valued random variable  $X$ ,

$$E[X] = - \int_{-\infty}^0 F(x) dx + \int_0^{+\infty} (1 - F(x)) dx.$$

Thus,

$$\begin{aligned} E[\tilde{v}_A(x, y)] - E[\tilde{v}_B(z, w)] = & \\ - \int_{-\infty}^0 G_A(x, y)(v) dv + \int_0^{\infty} (1 - G_A(x, y)(v)) dv + \int_{-\infty}^0 G_B(z, w)(v) dv - \int_0^{\infty} (1 - G_B(z, w)(v)) dv = & \\ & \int_{-\infty}^{+\infty} [G_B(z, w)(v) - G_A(x, y)(v)] dv. \end{aligned}$$

$\square$

*Proof of Theorem 2.* The proof is analogous to that of Theorem 1 in Alós-Ferrer et al. (2021), slightly condensed thanks to the use of the lemmata above. Consider any 2RUM-CF  $(u_A, \tilde{v}_A; u_B, \tilde{v}_B; r)$  that rationalizes  $(p_A, f_A; p_B, f_B)$ .

By Lemma 2,

$$\begin{aligned}
[u_A(x) - u_A(y)] - [u_B(z) - u_B(w)] &= E[\tilde{v}_A(x, y)] - E[\tilde{v}_B(z, w)] = \\
&= \int_{-\infty}^{+\infty} [G_B(z, w)(v) - G_A(x, y)(v)] dv = \\
\int_{-\infty}^0 [G_B(z, w)(v) - G_A(x, y)(v)] dv &+ \int_0^{+\infty} [G_B(z, w)(v) - G_A(x, y)(v)] dv = \\
\int_0^{+\infty} [G_B(z, w)(-v) - G_A(x, y)(-v)] dv &+ \int_0^{+\infty} [G_B(z, w)(v) - G_A(x, y)(v)] dv = \\
\int_0^{+\infty} [G_B(z, w)(v) + G_B(z, w)(-v) - G_A(x, y)(v) - G_A(x, y)(-v)] dv
\end{aligned}$$

For any  $v > 0$ , let  $t = r(v)$ . By Lemma 1,  $H_A(x, y)(t) \leq H_B(z, w)(t)$  can be rewritten as

$$G_B(v) + G_B(-v) \geq G_A(v) + G_A(-v)$$

for any  $v$  with  $t = r(v) > 0$ . The inequality follows for  $v = 0$  by continuity. For any  $v$  with  $r(v) = 0$ , it follows because in that case  $G_A(x, y)(v) = G_B(z, w) = 1$  and  $G_A(x, y)(-v) = G_B(z, w)(-v) = 0$ , as otherwise the corresponding RUM-CF would generate an atom at the response time of zero. We conclude that the inequality holds for all  $v \geq 0$ . Thus, the integral above is positive and the conclusion follows.

If  $H_A(x, y)(t) > H_B(z, w)(t)$  for some  $t$ , it is strict for a nonempty interval by continuity, and it follows that the final inequality is also strict.  $\square$

*Proof of Theorem 3.* “If.” Suppose (a) and (b) hold, and let  $(u_A, \tilde{v}_A; u_B, \tilde{v}_B; r)$  be a 2RUM-CF that rationalizes the 2SCF-RT. By (a),

$$p_A(y, x)F_A(y, x)(t) \leq p_B(w, z)F_B(w, z)(t) \quad \text{for all } t > 0,$$

which, by Definition 7, implies

$$G_A(x, y)(-r^{-1}(t)) \leq G_B(z, w)(-r^{-1}(t)).$$

This implies that  $G_A(x, y)(v) \leq G_B(z, w)(v)$  for any  $v < 0$  such that  $r(-v) = t > 0$ . For  $v = 0$ , by continuity, the inequality also holds. If  $r(v) = 0$  (as in the proof of Theorem 2), the inequality also holds since the absence of atoms at a response time of zero implies that  $G_A(x, y)(v) = 0$  and  $G_B(z, w)(v) = 1$ .

An analogous argument shows that  $G_A(x, y)(v) \leq G_B(z, w)(v)$  also for any  $v \geq 0$ . This completes the proof of the “if” direction.

“Only if.” suppose that the 2SCF-RT reveals that group  $A$  prefers  $x$  over  $y$  uniformly more than group  $B$  prefers  $z$  over  $w$ . Let  $(u_A, \tilde{v}_A; u_B, \tilde{v}_B; r)$  be a 2RUM-CF that rationalizes the 2SCF-RT. Hence,  $G_A(x, y)(v) \leq G_B(z, w)(v)$  for all  $v$ .

Analogously to the “if” direction, evaluating this inequality for  $-r^{-1}(t)$  proves (a), and evaluating it for  $r^{-1}(t)$  proves (b).  $\square$

*Proof of Corollary 2.* The argument is identical to the “only if” direction in the proof of Theorem 3, with the only difference that the 2RUM-CF is fixed and it rationalizes the 2SCF-RT because the latter is taken to be generated by the former.  $\square$

## Appendix B Additional Details on Survey Design

The sample size and power analysis is based on the tests of proportions for comparisons across groups. Given that in some cases participants might not be equally distributed between groups (e.g., political attitudes), we conservatively set the sample size to be able to allow for a 0.25 allocation ratio (80% vs. 20%). We further require to have enough power (0.8) to detect a 0.1 difference in the proportion of people supporting one option compared to the other between groups. The resulting sample size fulfilling these conditions is  $N = 1008$ . At the time the survey was conducted, a representative sample in Prolific required  $N = 1200$ , and hence we used this sample size.

During the survey, we implemented control questions for attention, recruiting participants who successfully passed the control questions until we reached the required sample size. Due to simultaneous online recruitment, the target was slightly exceeded ( $N = 1202$ ).

## Appendix C List of Questions

The actual order of questions in the survey was randomized.

- Group division based on gender (classification on the basis of gender reported in the prolific registration).

The first three questions are inspired by Bursztyn et al. (2020).

1. Are you in favor of gender quotas? [Yes/No]
2. Do you think women should earn less than men for the same job? [Yes/No]
3. Do you think mothers should stay at home with their kids instead of working? [Yes/No]
4. Would you cheat on a partner if given the occasion (and she/he would never find out)? [Yes/No]
5. Would you cheat on taxes if you had a chance? [Yes/No]

- Group division based on age (median split on the basis of age reported in the prolific registration).

The first two questions are inspired by Bechtel and Scheve (2013) and Bechtel et al. (2019). The third question in this group is inspired by Bechtel et al. (2020).

1. Do you think rich countries should pay more than poor countries to finance policies decreasing greenhouse gas emissions (independently of the individual history of emissions)? [Yes/No]
2. Do you think policies to decrease greenhouse gas emissions should be financed by rich countries only, instead of by all countries proportionally to current emissions? [Yes, only rich countries / No, proportionally to current emissions]
3. Do you think policies to decrease greenhouse gas emissions should impose increasing costs over time (countries pay more in the future than now) compared to constant costs? [Yes, increasing costs / No, constant costs over time]
4. Should the government impose a tax on industry to discourage industry practices that contribute to global warming? [Yes/No]



5. Are you in favor of legalising the use and consumption of Cannabis (marijuana)? [Yes/No]

- Group division based on political orientation (classification on the basis of question at the end, see below).

The first question is inspired by Frank (2016) while the second and third questions are inspired by Sides and Citrin (2007). The fourth question is inspired by Goddard and Smith (2001).

1. Do you think people get ahead mostly because of their own merits, or rather because of luck and help from others? [Yes, merit / No, luck and help]
2. Should your country allow more immigrants to come and live in it? [Yes/No]
3. Do you think your country should limit international trade to protect national jobs? [Yes/No]
4. Do you think people who don't work should have guaranteed access to health services? [Yes/No]
5. Do you think BREXIT was a good idea? [Yes/No]
6. Are you in favor of same-sex marriage? [Yes/No]
7. Do you think vaccination against COVID-19 should be mandatory? [Yes/No]

- Group division based on household income (on the basis of question at the end, see below).

The first two questions are inspired by Fehr et al. (2022). The third to sixth questions are inspired by Bechtel and Liesch (2020).

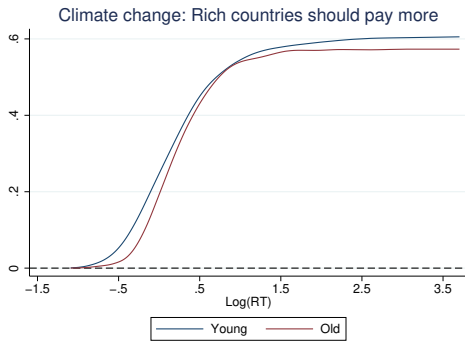
1. Are you in favor of sharply increasing taxation for people at the upper 1% of the income range? [Yes/No]
2. Are you in favor of a high inheritance tax? [Yes/No]
3. Would you support a policy which increases the average yearly income in your country by £5,000, but decreases your personal income by £2,500? [Yes/No]
4. Would you support a policy which increases your personal income by £2,500, but decreases the average yearly income in your country by £5,000? [Yes/No]
5. Would you support a policy which increases the average income of the lowest-income quarter of the population by £5,000, but decreases the average income of the rest of the population by £2,500? [Yes/No] (*Comparison for this question is lowest quarter vs. the rest*)
6. Do you think that it is the responsibility of the government to reduce the differences in income between people with high income and those with low income? [Yes/No]

Final questions for defining groups.

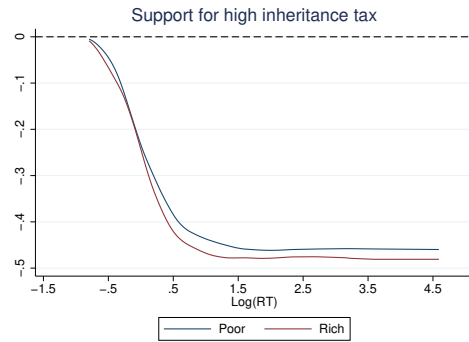
1. What is your household's approximate annual income? [less than £18,000][£18,000; £29,900][£29,901; £62,000][more than £62,000]
2. What describes you best politically: leaning more toward the left or toward the right? [Left/Right]

# Appendix D Complete Survey Results

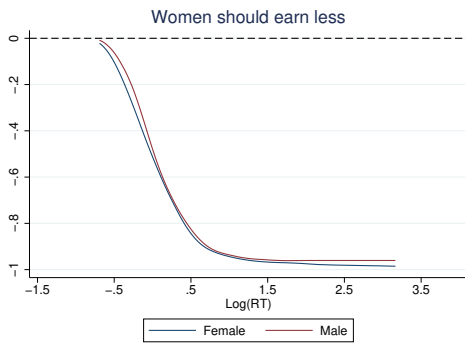
## D.1 Getting More from the Data



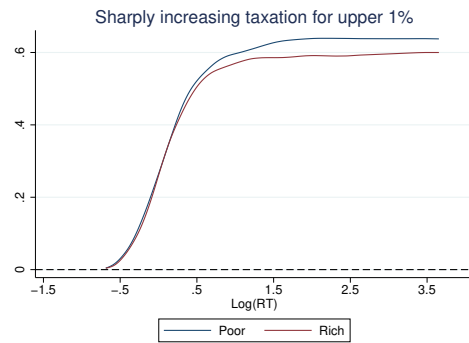
Old: 78.67% vs. Young: 80.40%,  
 $z = -0.744, p = 0.457$ .



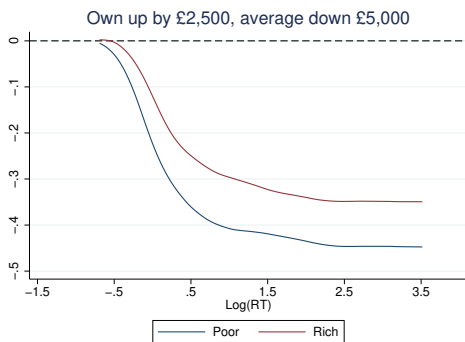
Rich: 25.96% vs. Poor: 26.98%,  
 $z = -0.368, p = 0.713$ .



Female: 6.88% vs. Male: 19.75%,  
 $z = -1.902, p = 0.057$ .

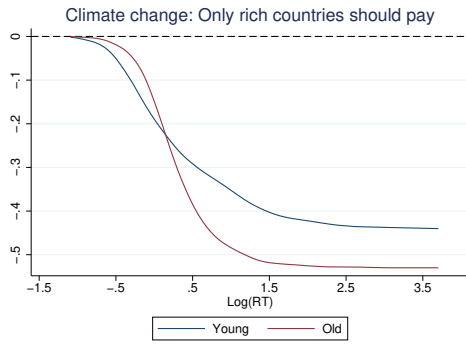


Rich: 80.00% vs. Poor: 81.84%,  
 $z = -0.743, p = 0.458$ .

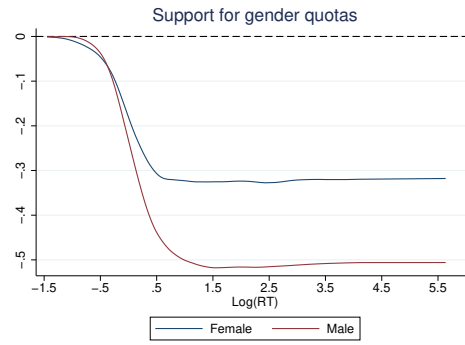


Rich: 32.60% vs. Poor: 27.63%,  
 $z = 1.734, p = 0.083$ .

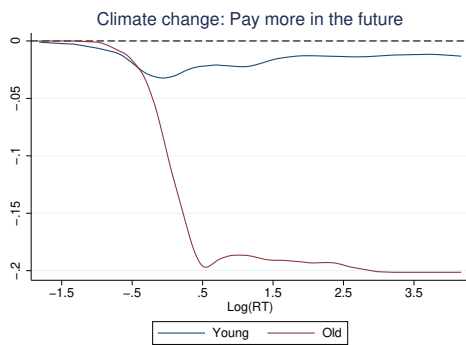
## D.2 Unwarranted Conclusions



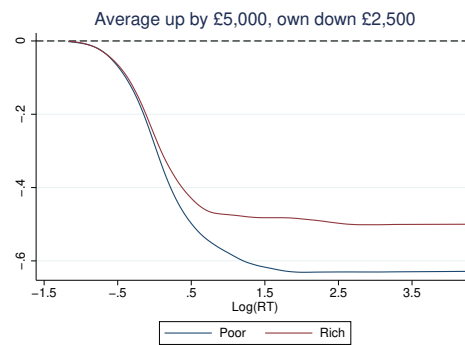
Old: 23.50% vs. Young: 27.91%,  
 $z = -1.748, p = 0.040$ .



Female: 34.15% vs. Male: 24.70%,  
 $z = 3.588, p < 0.001$ .

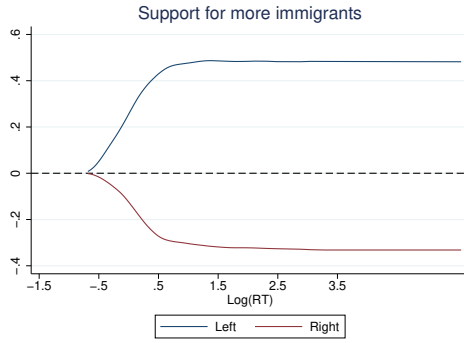


Old: 39.93% vs. Young: 49.28%,  
 $z = -3.256, p = 0.001$ .

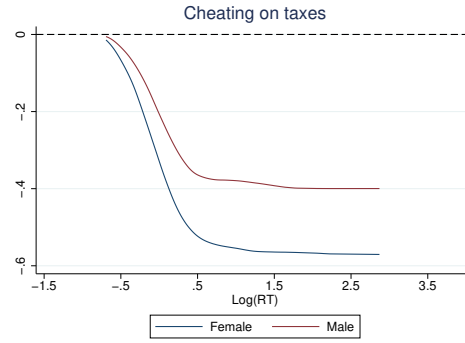


Rich: 25.00% vs. Poor: 18.60%,  
 $z = 2.537, p = 0.011$ .

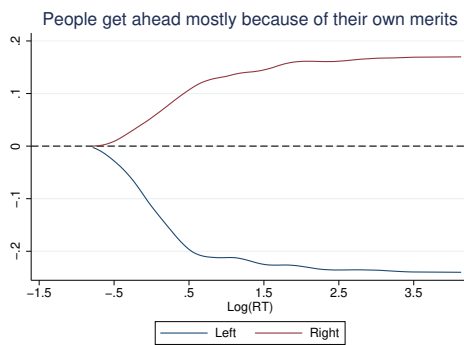
### D.3 Positive Concordance



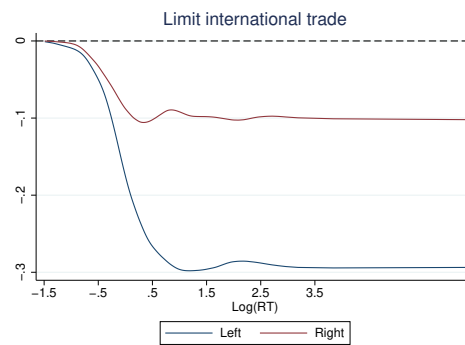
Right: 33.41% vs. Left: 74.06%,  
 $z = -13.488, p < 0.001$ .



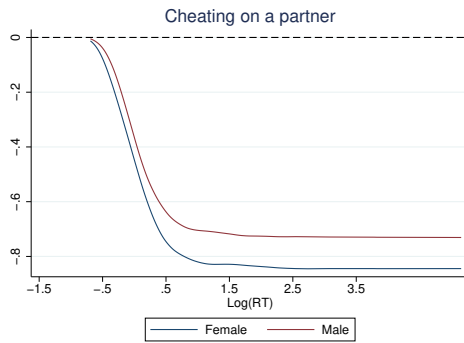
Female: 21.43% vs. Male: 30.02%,  
 $z = -3.315, p < 0.001$ .



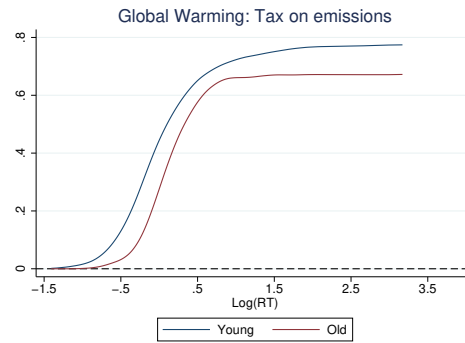
Right: 58.49% vs. Left: 37.97%,  
 $z = 6.838, p < 0.001$ .



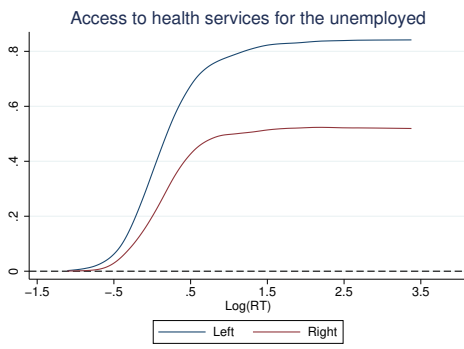
Right: 44.90% vs. Left: 35.35%,  
 $z = 3.274, p = 0.001$ .



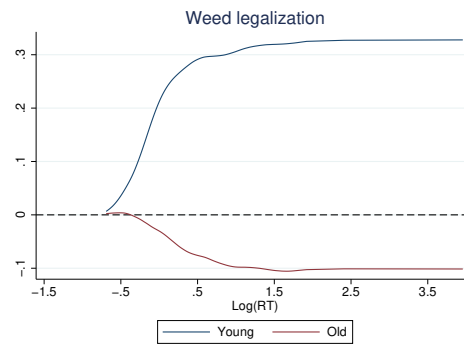
Female: 7.82% vs. Male: 13.46%,  
 $z = -3.118, p = 0.002$ .



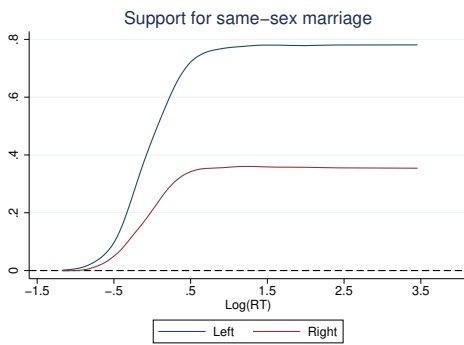
Old: 83.65% vs. Young: 88.73%,  
 $z = -2.557, p = 0.005$ .



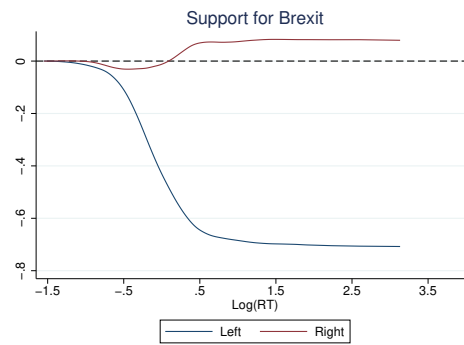
Right: 75.96% vs. Left: 92.21%,  
 $z = -7.823, p < 0.001$ .



Old: 44.93% vs. Young: 66.43%,  
 $z = -7.331, p < 0.001$ .

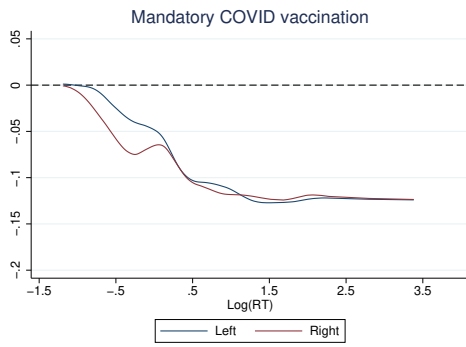


Right: 67.65% vs. Left: 89.05%,  
 $z = -9.131, p < 0.001$ .

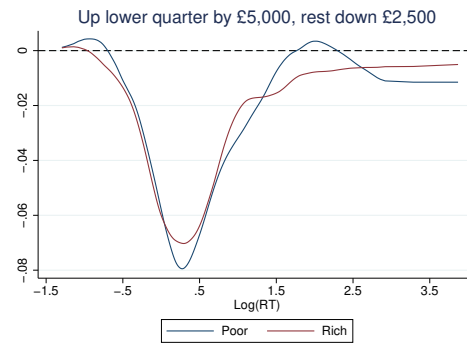


Right: 53.97% vs. Left: 14.59%,  
 $z = 14.497, p < 0.001$ .

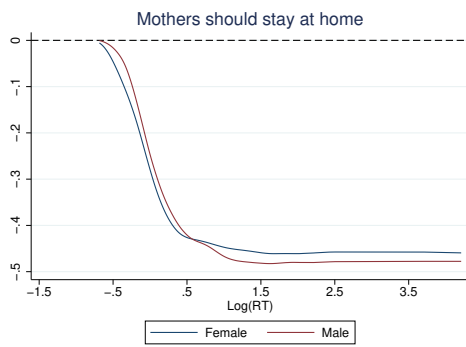
## D.4 Negative Concordance



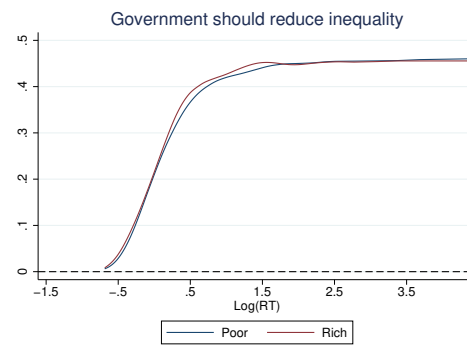
Right: 43.76% vs. Left: 43.80%,  
 $z = -0.012, p = 0.991$ .



Rich: 49.78% vs. Poor: 49.43%,  
 $z = 0.085, p = 0.933$ .



Female: 26.99% vs. Male: 26.12%,  
 $z = 0.3327, p = 0.739$ .



Rich: 72.78% vs. Poor: 73.03%,  
 $z = -0.091, p = 0.928$ .