ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Hufe, Paul; Weishaar, Daniel

Working Paper

Just Cheap Talk? Investigating Fairness Preferences in Hypothetical Scenarios

CESifo Working Paper, No. 11647

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Hufe, Paul; Weishaar, Daniel (2025) : Just Cheap Talk? Investigating Fairness Preferences in Hypothetical Scenarios, CESifo Working Paper, No. 11647, CESifo GmbH, Munich

This Version is available at:
https://hdl.handle.net/10419/314686

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.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

CESifo WORKING PAPERS

11647
2025

January 2025

Just Cheap Talk? Investigating Fairness Preferences in Hypothetical Scenarios

Paul Hufe, Daniel Weishaar

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 11647*

Just Cheap Talk? Investigating Fairness Preferences in Hypothetical Scenarios

Abstract

The measurement of preferences often relies on surveys in which individuals evaluate hypothetical scenarios. This paper proposes and validates a novel factorial survey tool to measure fairness preferences. We examine whether a non-incentivized survey captures the same distributional preferences as an impartial spectator design, where choices may apply to a real person. In contrast to prior studies, our design involves high stakes, with respondents determining a real person's monthly earnings, ranging from $500 to $5,700. We find that the non-incentivized survey module yields nearly identical results compared to the incentivized experiment and recovers fairness preferences that are stable over time. Furthermore, we show that most respondents adopt intermediate fairness positions, with fewer exhibiting strictly egalitarian or libertarian preferences. In sum, these findings suggest that high-stake incentives do not significantly impact the measurement of fairness preferences and that non-incentivized survey questions covering realistic scenarios offer valuable insights into the nature of these preferences.

JEL-Codes: C900, D630, I390.

Keywords: fairness preferences, survey experiment, vignette studies.

Paul Hufe
University of Bristol / United Kingdom
paul.hufe@bristol.ac.uk

Daniel Weishaar
LMU Munich / Germany
daniel.weishaar@econ.lmu.de

January 20, 2025
This paper benefited strongly from discussions with Ingvild Almås. It is part of a larger research project in which we measure fairness preferences and beliefs about inequality around the world (see Riksbankens Jubileumsfond (P22-0564) and UKRI Future Leaders Fellowship (MR/X033333/1): "(Un)Fair inequality in the labor market: a global study"). We received useful feedback from seminar audiences at LMU Munich. This research was funded by the British Academy (TDA21/210082) and by Deutsche Forschungsgemeinschaft through CRC TRR 190 (No. 280092119). Hufe gratefully acknowledges financial support from UKRI (MR/X033333/1). Weishaar gratefully acknowledges financial support from the Joachim Herz Foundation and the Fritz Thyssen Foundation. The questionnaire and core analysis were pre-registered via the Open Science Framework (OSF), No. DV3KP. We obtained ethical approval from the Institutional Review Boards at the University of Bristol, LMU Munich, and NHH Norwegian School of Economics. All remaining errors are our own.

1 Introduction

There is expanding literature in economics and other social sciences that investigates which inequalities are seen as unfair by people (Alesina et al., 2018; Almås et al., 2020; Andre, forthcoming; Cappelen et al., 2007; Jasso and Webster, 1999; Konow, 2000). In these papers, fairness preferences are typically elicited using incentivized experiments or non-incentivized surveys. Researchers considering the choice between these two research designs face a trade-off. On the one hand, experiments combine stylized representations of real-world situations with payout-relevant choices of respondents. On the other hand, survey questions often mirror real-world contexts more closely; however, respondents' answers have no consequences in the real world. Therefore, survey-based methods are often considered unreliable predictors of actual behavior. This raises the question of whether researchers can employ non-incentivized surveys to analyze fairness preferences or whether such answers must be considered "cheap talk." In this paper, we address this question by using a representative sample of the US adult population to test whether answers to hypothetical questions align with those from an incentivized experiment.

Our survey tool integrates core functionalities of impartial spectator experiments (Almås et al., 2024a; Almås et al., 2020; Andre, forthcoming; Cappelen et al., 2013; Konow, 2000; Konow et al., 2020) with the methodological advantages of factorial surveys (Auspurg et al., 2017; Gaertner and Schwettmann, 2007; Jasso and Webster, 1999; Konow, 1996).[1] The questions in our survey tool show respondents pairs of hypothetical persons that are described in terms of observable characteristics, i.e., their gender, age, educational attainment, parental background, working hours, and labor market earnings. Respondents then distribute earnings between these two persons based on the given information.

In impartial spectator designs, choices are not payoff-relevant for the respondents themselves. Yet, answers in a hypothetical task might still differ from their incentivized analogs. First, hypothetical distribution tasks are prone to different biases, including experimenter demand effects and social desirability biases (Stantcheva, 2023). For instance, in the absence of real-world consequences, respondents may tend towards more equal income allocations to comply with perceived expectations from the researcher. Such demand effects are less of a concern when real money is at stake (Haaland et al., 2023). Second, even without such systematic biases, hypothetical distribution tasks may be prone to measurement error, e.g., if non-incentivized respondents are less attentive when completing the tasks.

Our survey tool differs from prior literature on fairness preferences which has largely focused on the extent to which individuals reward rather abstract concepts such as "luck," "productivity," "hard work," and "talent" (e.g., Cappelen et al., 2010; Mollerstrom et al., 2015). In contrast to these studies, our survey tool allows us to elicit fairness preferences that can be directly mapped to observable labor market inequalities, e.g., gender gaps (Blau and Kahn, 2017),

[1] For detailed reviews on experimental and survey-based evidence on fairness preferences, see Almås et al. (2023), Almås et al. (2024b), and Gaertner and Schokkaert (2012).

returns to hours (Kuhn and Lozano, 2008), education premia (Harmon et al., 2003), and intergenerational persistence (Roemer and Trannoy, 2016). However, the focus on these real-world inequalities also makes it prohibitively costly to elicit the relevant preferences in an experimental design where respondents' choices are consequential for the earnings of actual persons. Therefore, it is important to test whether hypothetical distribution tasks deliver credible results that align with the "gold standard" of incentivized experiments.

To address this question, we collected data from 1,602 adults in the United States between October and November 2022. The sampling was designed to be representative of various demographic characteristics such as gender, age, education, employment status, and region of residence. The survey modules, as well as core analyses, were pre-registered via the Open Science Framework (OSF), No. DV3KP.

We validate our survey tool along three dimensions. First, we test whether the distributional choices of respondents are different if they are payoff-relevant. For this purpose, we run an experiment with a between-subject design. All respondents answer a survey where they face a selection of tasks from our survey tool. Respondents in the treatment group are informed that one of the persons shown to them is a real person and that the decision made by a randomly chosen respondent will determine the monthly earnings of this individual. Thus, in contrast to the control group, they know that each choice may have substantial financial consequences for a real person. This design allows us to test whether fairness preferences in our hypothetical tool are consistent with the "gold standard" of an incentivized experiment (Bauer et al., 2020; Enke et al., 2022; Falk et al., 2023). Second, we test whether the distributional choices are stable over time. For this purpose, we employ a within-subject design and run an obfuscated follow-up one week after the baseline survey.[2] In particular, we invite respondents to another survey, where they again face a selection of tasks from our survey tool. Some tasks are repeated from the baseline wave, allowing us to calculate intertemporal correlations. This design enables us to test the stability of fairness preferences in our hypothetical tool and gives crucial information on measurement error in the elicited preference data (Gillen et al., 2019; Stantcheva, 2023). Lastly, next to the methodological validation of the survey design, we conduct a suggestive substantive analysis. Specifically, we describe the nature of fairness preferences identified through our survey and the heterogeneity of fairness views within the US population. This analysis has several caveats since the survey design was premised on methodological validation. Nevertheless, the substantive analysis provides an important cross-check on whether our hypothetical survey tool recovers preferences that are consistent with previous studies on fairness preferences in the US (Almås et al., 2020; Fisman et al., 2023; Konow et al., 2020).

Our results can be summarized as follows. First, the distributional choices of respondents are not affected by making them relevant to the earnings of real persons. The point estimates for

[2]This choice of the time lag is consistent with several recent articles that validate survey-based measurement tools using time-lags between one and two weeks (Bauer et al., 2020; Enke et al., 2022; Falk et al., 2023; Fallucchi et al., 2020).

the "real-person treatment" are small and insignificant at conventional levels of statistical significance. This conclusion remains unaffected when considering treatment effects on the distribution of choices and treatment effects within various population subgroups. Second, the distributional choices of respondents are relatively stable over time. The average (intra-respondent) intertemporal correlation of distributional choices is 0.56, which lies in the range of test-retest correlations of other preference survey modules (Enke et al., 2022). Furthermore, the intertemporal correlation is slightly higher in the incentivized group, suggesting that incentives have a small positive effect on reducing measurement error in the elicited preferences. Third, we find that the nature of the recovered fairness preferences is broadly consistent with previous studies on the US. Inequality acceptance ranges between Gini coefficients of 0.30 and 0.53 (e.g., Almås et al., 2020), the majority of respondents adopt intermediate fairness positions that are influenced by discretionary variables such as education and working hours (e.g., Konow, 2000), and the distributional choices of different population subgroups are consistent with self-serving biases (Costa-Font and Cowell, 2014). In summary, the results from our validation suggest that the proposed hypothetical survey tool recovers fairness preferences that are consistent with incentivized choices, stable, and reasonable in light of the existing literature.

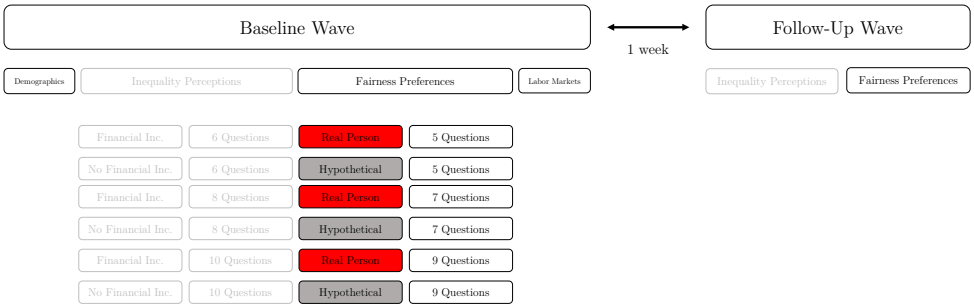
This paper contributes to two strands of the literature. First, we contribute to the literature on fairness preferences. There is a large literature in economics and other social sciences trying to understand the nature and anatomy of fairness preferences in different population groups (Almås et al., 2020; Andre, forthcoming; Cappelen et al., 2007; Gaertner and Schokkaert, 2012; Jasso and Webster, 1999; Konow, 2000; Starmans et al., 2017). In this paper, we validate a vignette-based survey tool that allows researchers to investigate fairness preferences in a flexible and cost-efficient way. Therefore, this study provides a crucial step to strengthen the methodological toolkit for investigating fairness preferences in applied research. Furthermore, there is a growing literature investigating the upstream determinants and downstream consequences of fairness preferences (Adriaans, 2023; Alesina et al., 2018; Andersen et al., 2023; Fehr et al., 2024). These studies often rely on survey-based measures of fairness preferences. Our results provide encouraging news for such research designs as the consistency of hypothetical and incentivized choices suggests that survey-based measures are not systematically biased compared to their incentivized analogs. Second, we contribute to a growing methodological literature that validates survey-based measurement tools in various domains, including risk, time, competition, and social preferences (Bauer et al., 2020; Enke et al., 2022; Falk et al., 2023; Fallucchi et al., 2020). To the best of our knowledge, our study is the first validation of an impartial spectator task under high monetary stakes.

The remainder of the paper is organized as follows. Section 2 describes the survey tool and provides information on the data collection. In Section 3, we present results for the effects of the "real-person treatment." Section 4 describes the stability of fairness preferences. Section 5 provides a suggestive comparison of the recovered fairness preferences relative to existing literature. Section 6 concludes the paper.

2 Survey Tool and Data Collection

Survey structure. Figure 1 provides an overview of the survey used in our analysis. The survey is structured into two waves, each consisting of multiple modules. In the first module of the baseline wave, we elicit the demographic characteristics of respondents. The second and third modules measure inequality perceptions and fairness preferences. The final module contains additional questions about the labor market. The two modules of the follow-up wave mirror the perceptions and preference modules of the baseline wave. In this paper, we focus exclusively on fairness preferences.[3]

Figure 1: Survey Structure



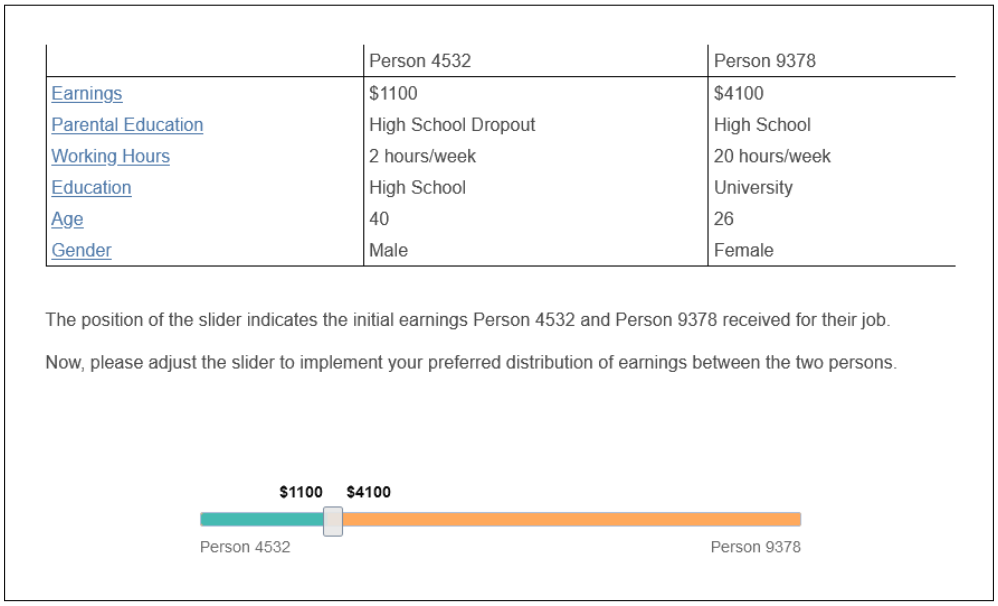
Note: This figure visualizes the structure of the survey with two waves (baseline, follow-up). Each wave consists of multiple modules. The modules on inequality perceptions are blurred out since they are not covered in this paper. The main treatment group (control group) is highlighted in red (gray).

Preference module. The module on fairness preferences consists of multiple questions that follow the design of factorial survey experiments. Factorial survey experiments are well-established tools in the social sciences to assess preferences and beliefs (Auspurg et al., 2015; Auspurg et al., 2017; Fisman et al., 2020; Jasso and Webster, 1997; Jasso and Webster, 1999; Wiswall and Zafar, 2018). In such experiments, respondents evaluate multiple hypothetical scenarios that vary at random in pre-defined characteristics. The random variation of characteristics has two main advantages. First, the design can replicate the complexities of the real world. In particular, respondents are forced to make trade-offs and weigh the importance of different real-world attributes against each other when making their choices. Second, the simultaneous variation of characteristics mitigates experimenter demand effects and social desirability biases—concerns that are particularly relevant in the domain of fairness preferences (Auspurg and Hinz, 2014).

In our survey module, respondents receive information on the observational characteristics of two persons—see Figure 2 for an example. We describe these persons in terms of six characteristics, i.e., their gender, age, own education, parental education, working hours, and labor market earnings. The choice of these characteristics was guided by the following

[3]The perceptions modules are designed to assess respondents' perceptions of inequality in the labor market. All treatments in this module are independent of the treatments in the preference module, allowing us to analyze these data in isolation.

Figure 2: Fairness Preference Elicitation, Exemplary Question



Note: This figure provides an example of a question screen in the fairness preference module. Each question shows the characteristics of two persons in six dimensions (earnings, gender, age, education, parental education, working hours) in a table format. Each of the two persons has been allocated a random person identifier from 1 to 9999. The ordering of characteristics in the table is randomized at the respondent-question level. A slider allows respondents to select their preferred distribution of earnings between the two persons. The chosen allocation is also shown numerically above the slider.

considerations. First, we selected the dimensionality of the vignettes by acknowledging the trade-off between precision and complexity in the design of factorial experiments. On the one hand, a higher number of vignette characteristics has the benefit that more confounding factors can be held constant. On the other hand, increasing the number of characteristics also puts a larger cognitive burden on participants when making distributional choices. With six dimensions, we opted for an intermediate level of complexity (Auspurg and Hinz, 2014). Second, we then selected characteristics that are both relevant for understanding earnings inequality and widely available in standard household survey data (Bick et al., 2022; Goldin, 2014; Lemieux, 2006; Magnac and Roux, 2021; Mazumder, 2005). We are aware that the resulting selection of characteristics is somewhat discretionary. Therefore, we also validate our selection ex-post by asking respondents about characteristics they consider particularly important when making distributional choices. Appendix Figure A1 shows that five out of the six selected characteristics are chosen by at least ten percent of respondents. Furthermore, many of our chosen characteristics correlate strongly with other characteristics that participants find important, e.g., age (our selection) and work experience (respondents selection).

Each characteristic can take multiple expressions. For instance, the characteristic of education can take three values, i.e., *High School Dropout*, *High School*, or *University*. The potential expressions for each characteristic are shown in Table 1. The order of characteristics is randomized at the respondent-question level to ensure that results are not driven by order

effects (Day et al., 2012).

Table 1: Fairness Preference Elicitation, Characteristics and Expressions

Characteristic	Number	Displayed Values for Expression
Gender	2	Male / Female
Age	2	(26 , 35 , 40) / (50, 55, 59)
Education	3	High School Drop-Out / High School / University
Parental Education	3	High School Drop-Out / High School / University
Working Hours	4	(2, 5) / (20, 27, 31) / (39, 40) / (48, 51, 59)
Earnings	5	$1,100 / $2,700 / $4,100 / $5,900 / $11,400

Note: This table shows the characteristics displayed in questions of the fairness preference module (column 1), the number of coarse expressions within each characteristic (column 2), and the displayed values for each expression (column 3). We employ a second randomization for age and working hours to display exact values instead of ranges. For each of the two age range groups (25-44, 45-65) and each of the four working hours range groups (1-9, 10-34, 35-44, More than 44), we draw from integers in the respective range.

Combining all potential expressions yields a set of 720 profiles ($2 \times 2 \times 3 \times 3 \times 4 \times 5$) and a set of 258,840 ($\frac{720 \times 719}{2}$) unique unordered profile pairs. In the following, we will refer to each unique profile as a *vignette person* and each unique profile pair as a *vignette*. We show respondents a selection of vignettes determined via a random draw from the full set. Based on the presented information, respondents can use a slider to adjust the initial earnings and to implement their preferred earnings distribution in the vignettes. All respondents answered the same selection of vignettes, with some variation in the number of vignettes per respondent—see also our discussion on the "length treatment" below.

Treatment 1: "Real-person treatment." To assess whether hypothetical questions in factorial surveys recover fairness preferences consistent with incentivized experiments, we randomize respondents into two groups. Respondents in the control group complete a series of hypothetical distribution tasks. Respondents in the treatment group complete the same series of tasks. However, they are informed that one of the vignette persons is a real person. Furthermore, they are informed that the choice of one respondent will be selected to determine the monthly earnings of this person. Respondents know that the real person is between 25 and 65 years old, is a resident of the United States, and works in a job to earn money. Importantly, this information does not allow respondents to distinguish the real person from any other vignette person. They also know that the total earnings of the real person consist of two parts: (i) a fixed payment of $500 and (ii) a flexible payment that can be changed by the respondent.

The research team hired a real person in August 2022—see Appendix Table A1 for the characteristics of the real person as displayed in their vignette. The hired person was informed that the fixed-term contract would have a duration of one month and that the exact amount of their earnings would be determined by another individual; however, they did not know the exact process of how this happens. The person only knew their total earnings would be $500 or above. To determine the potential earnings of the real person, we proceeded in two steps. First, we allocated the real person a monthly earnings value by randomly drawing from

the set of potential earnings displayed in Table 1. Second, we randomly matched the real person with another (hypothetical) vignette person. This two-step procedure fixed the volume of earnings in the vignette of the real person at \$5,200. Therefore, including the fixed payment of \$500, the upper bound of potential earnings for the real person was \$5,700. This upper bound would be realized if the decisive respondent allocated all the vignette earnings to the real person. Importantly, the vignette with the real person was presented alongside all other vignettes, and the identity of the real person was concealed from respondents. Consequently, respondents also faced situations in which the potential earnings implications were even higher. The average earnings volume in the displayed vignettes—and therefore the average upper bound of potential payments to the real person from the respondents' perspective—was \$11,420 in the main vignettes of the preference module. Furthermore, we ensured the salience of the real-world consequences through a training task. Specifically, we trained respondents on an example vignette and highlighted the potential earnings consequences for the real person after respondents had made their distributional choices.

Our incentivization is based on high stakes with a low probability of implementation. Alternatively, we could have used an incentive scheme based on lower stakes but a higher probability of implementation. Due to our focus on real-world characteristics, the former path appears more natural since it allows us to focus on payouts that reflect realistic monthly wages drawn from the US working population. However, we contend that a generalization of our results to alternative incentive schemes is an interesting path for future research.

Treatment 2: Length treatment. To assess the sensitivity of our conclusions to the length of the survey, we vary the number of vignettes in the survey module. All respondents made at least five distributional choices; however, 1/3 of respondents received 2 or 4 additional vignettes, respectively. For the main validation, we focus on the first five questions answered by all respondents and use the variation from the length treatments in robustness analyses. The assignment to the length treatments was independent of the allocation to the "real-person treatment." Therefore, we obtain six groups of approximately equal size that vary in their exposure to the "real-person treatment" and the survey length (Figure 1).

Baseline wave. We administered the baseline wave of the survey to 1,602 adult citizens of the United States. Data collection took place between October and November 2022. Respondents were contacted through the survey provider *Dynata* and received a participation payment depending on the expected survey length.[4] The mean (median) completion time was 24 (19) minutes for the baseline wave (Appendix Figure A2).

Respondents were targeted to match the population along five dimensions (gender, age, education, employment status, and region of residence). In Panel A of Table 2, we compare

[4]For the baseline wave, respondents were able to earn between \$0.20 and \$1.50. For the shorter follow-up wave, the payment varied between \$0.10 and \$1.20. The varying participation payment was used by *Dynata* to obtain responses from demographic segments of the population that are more difficult to reach.

our sample to the American Community Survey (ACS) regarding the targeted characteristics. In general, we match the data well. Our sample has a slight underrepresentation of people with low education. In addition, we received over-proportional (under-proportional) responses from mid-western states (southern states). Panel B of Table 2 further shows that our sample is also broadly representative regarding other observable characteristics like ethnicity and income.

We take various steps to ensure the quality of survey answers. First, we included an attention check at the beginning of the preference module. Respondents who failed this attention check were screened out directly and were not part of the sample. Second, we asked a training question after explaining the tasks in the preference module. Around 72% of respondents passed this question on the first try. In robustness checks, we show that our results are not sensitive to excluding respondents who did not pass the training question on their first attempt.

Follow-up wave. We invited respondents to a follow-up wave one week after they completed the baseline wave. Among others, the follow-up wave consists of a fairness preference module with six questions. As in the baseline wave, respondents faced a selection of distribution tasks based on the vignettes from our survey module. Three of the questions are repetitions from the baseline wave, which were presented to the respondents in an obfuscated way. This feature allows us to assess the stability of fairness preferences over time.

The follow-up obtained a response rate of around 44%, and around 90% of respondents answered within two weeks. The resulting sample is slightly older but otherwise broadly comparable to our baseline sample regarding observable demographics (Appendix Table A2). The mean (median) completion for the follow-up wave was 14 (9) minutes (Appendix Figure A2).

3 Effects of the "Real-Person Treatment"

In this section, we investigate whether the potential for real-world implementation affected the distributional choices of respondents. First, we present methodological checks on the randomization and the anonymity of the real person. Second, we present the treatment effects on distributional choices. Third, we investigate potential heterogeneities by population subgroups. Fourth, we present robustness analyses. All analyses in this section are pre-registered unless noted otherwise.

Balancing. To give our estimates a causal interpretation, the treatment assignment must be uncorrelated with any respondent characteristics that may predict their distributional choices. Therefore, we test the balance of respondents' socio-demographic characteristics

Table 2: Demographics, Comparison to ACS

Panel A: Hard Quota Demographics

Variable	Survey (%)	ACS (%)
Gender (N=1589)		
Male	48.58	48.66
Female	51.42	51.34
Age (N=1602)		
18 - 24	12.98	11.89
25 - 34	13.98	17.85
35 - 44	16.29	16.48
45 - 54	15.61	15.97
55 - 64	18.04	16.63
Older than 64	23.10	21.18
Education (N=1602)		
Lower Education	4.31	11.43
Middle Education	39.95	27.58
Higher Education	55.74	60.99
Employment Status (N=1602)		
In Labor Force	61.92	61.97
Unemployed	5.74	2.83
Not in Labor Force	32.33	35.19
Region (N=1602)		
North-East	19.35	17.43
Mid-West	33.52	20.76
West	23.91	23.76
South	23.22	38.05

Panel B: Other Selected Demographics

Variable	Survey (%)	ACS (%)
Ethnicity (N=1598)		
White / Caucasian	81.29	73.60
Black / African American	9.45	12.47
American Indian / Alaska Native	1.31	0.82
Asian / Asian American / Native Hawaiian / Pacific Islander	4.07	6.08
Other	3.88	7.03
Hispanic (N=1600)		
Hispanic	10.06	16.40
Not Hispanic	89.94	83.60
Income (N=1599)		
Did not work	22.33	32.83
$0 - $25,000	19.45	24.06
$25,000 - $40,000	17.07	13.27
$40,000 - $59,000	14.20	9.94
$59,000 - $88,000	12.26	9.83
More than $88,000	14.70	10.00

Note: This table compares the sample of our baseline survey to the American Community Survey (ACS, 2019). Panel A (B) shows hard quota demographics (other selected demographics). Sample sizes vary across variables since we omit small answer categories in gender ("Non-binary"), ethnicity ("Prefer not to answer"), hispanic ("Prefer not to answer"), and income ("Prefer not to answer"). Appendix Table A2 presents the full set of demographics.
Source: Own calculations based on survey responses and American Community Survey (ACS, 2019).

between the treatment and the control group. In particular, we regress the treatment status $Treat_i$ of respondent i on K pre-specified individual characteristics denoted by x_i^k:

$$x_i^k = \alpha^k + \beta^k \ Treat_i + \varepsilon_i^k. \tag{1}$$

Table 3 presents the results. In this table, we show sample sizes, point estimates of β^k, mean outcomes of the control group, and p-values associated with β^k. We account for multiple hypothesis testing by correcting for the family-wise error rate using the step-down procedure of Romano and Wolf (2005) and Romano and Wolf (2016).

Table 3: Balancing Tests

Demographics	Binary Split	N	Point Estimate	Control Mean	RW p-values
Gender	Male vs. Female	1589	-0.007	0.5176	0.987
Age	$<$ 45 years vs. $\geq$ 45 years	1602	-0.050	0.5925	0.241
Ethnicity	White/Caucasian vs. Other	1598	0.006	0.1842	0.987
Education	$\leq$ HS vs. $>$ HS	1602	0.037	0.5387	0.543
Employment	(Self)Employed vs. Other	1602	-0.018	0.3900	0.938
Hours	$<$ 35 hours vs. $\geq$ 35 hours	1468	-0.019	0.5249	0.938
Income	$<$ \$40,000 vs. $\geq$ \$40,000	1599	-0.003	0.4128	0.987

Note: This table presents results of the balancing test outlined in equation (1). The table shows point estimates for the coefficient of interest, the mean of the control group (first group in binary split), and associated heteroskedasticity robust p-values adjusted for multiple hypothesis testing using the Romano and Wolf (2016) step-down procedure with 1000 bootstrap draws. Sample sizes differ across demographic characteristics due to the exclusion of responses such as "Prefer not to answer" (see Table 2).
Source: Own calculation based on survey responses.

For all considered characteristics, we find point estimates that are close to zero and small compared to the control mean. Furthermore, we cannot reject the null hypothesis of $\beta^k = 0$ at conventional levels of statistical significance. These results suggest that our randomization was successful and that we can give our treatment effects a causal interpretation.[5]

Anonymity of the real person. Out of all the vignettes faced by respondents, only one choice can determine the earnings of the real person. Respondents are not told which decision is hypothetical, and therefore, they are encouraged to treat each vignette as if it were payoff-relevant. The equal consideration of vignettes, however, would be threatened if respondents could identify the real person. In this case, the treatment would have no bite for vignettes that do not involve the real person (Andre, forthcoming).

To address this concern, we asked respondents to identify the real person at the end of the preference module. To facilitate this task, we allowed respondents to look up all previous vignettes. Furthermore, we strongly incentivized the identification of the real person by paying out \$5 for a correct answer. Despite these incentives, only about 5% of respondents

[5]We pre-specified a joint balancing test that included a treatment in the preference module and a treatment in the perceptions module. We depart from the pre-analysis plan since we focus on the preference module in this paper. However, the pre-specified joint balancing test leads to the same conclusion as the balancing test presented above (Appendix Table B3).

guessed correctly (Appendix Figure B3). This number is close to the number that would be obtained if all respondents were to guess by chance.[6] This result suggests that we successfully preserved the anonymity of the real person and supports the assumption that respondents considered all choices equally likely to be payoff-relevant.

Treatment effects on distributional choices. We estimate average treatment effects through ordinary least-squares using the following model:

$$\Delta y_{ij} = \alpha^j + \beta^j \, Treat_i + \mathbf{x}_i' \, \boldsymbol{\delta}^j + \varepsilon_{ij}, \tag{2}$$

where Δy_{ij} is the difference in money allocated to vignette persons A and B by respondent i in vignette j. $Treat_i$ is the binary treatment indicator. Per our pre-analysis plan, we also include the vector $\mathbf{x}_i'$ to control for demographic variables that are found to be unbalanced. Due to the successful randomization into treatment, this vector is empty (see Table 3).

Table 4 displays the results for each of the five vignettes that were answered by all respondents. We present point estimates of β^j, means of the outcome variables in the control group, and a set of three different p-values associated with β^j: (i) uncorrected analytical model p-values, (ii) p-values based on bootstrapped standard errors, (iii) p-values that account for multiple hypothesis testing through the step-down procedure of Romano and Wolf (2005) and Romano and Wolf (2016).

Table 4: Allocations, "Real-Person Treatment"

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	1602	114.52	1038.02	0.324	0.322	0.841
Q2	1602	8.37	-2298.01	0.931	0.932	0.932
Q3	1602	211.77	2036.38	0.591	0.598	0.841
Q4	1602	-256.17	-4072.08	0.321	0.325	0.841
Q5	1602	-264.02	6184.43	0.343	0.337	0.841

Note: This table presents results of the regression analysis outlined in equation (2). We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculation based on survey responses.

Point estimates of β^j are small and range between \$8 and \$264. This range corresponds to 0.4%–11.0% of the control mean. For none of the five considered vignettes, we can reject the null hypothesis that average choices are equal across treatment and control groups. Appendix Table B4 presents results from non-pre-registered analyses using alternative scalings of the outcome variable. In particular, we repeat the previous analysis by replacing Δy_{ij} with the

[6] Depending on the length treatment, the preference module contains five, seven, or nine questions. Therefore, if all respondents guessed randomly, we would expect that 7.6% ($= \frac{1}{3} \times [\frac{1}{10} + \frac{1}{14} + \frac{1}{18}]$) identify the real person correctly.

income share allocated to person A in the vignette. In this case, we find treatment effects that range between 0.1 and 1.1 percentage points. This corresponds to 0.29%–2.25% of the control mean without any significant differences between the treatment and control groups.

The average treatment effects may conceal offsetting treatment responses at different parts of the outcome distribution. To assuage this concern, we plot distributions of Δy_{ij} separately for treatment and control groups (Figure 3). The left panels show frequency distributions, whereas the right panels show corresponding cumulative distribution functions for each considered vignette. Visual inspection shows that the distributions largely overlap. This diagnosis is confirmed by Kolmogorov–Smirnov tests, according to which we cannot reject the null hypothesis that the implemented distributions are the same in treatment and control groups (Appendix Table B5). These results show that our null findings concerning average treatment effects are not the result of offsetting responses at different parts of the outcome distribution.

Heterogeneity analysis. Different population groups might react differently to the potential real-life consequences of a distributional choice. The absence of heterogeneous treatment effects, however, is important since many analysts might be interested in using hypothetical questions to assess differences in fairness preferences across population groups. If different population groups had a different propensity to reveal their preferences in hypothetical settings, analyses might erroneously detect group differences in fairness preferences, or the absence thereof, in such settings.

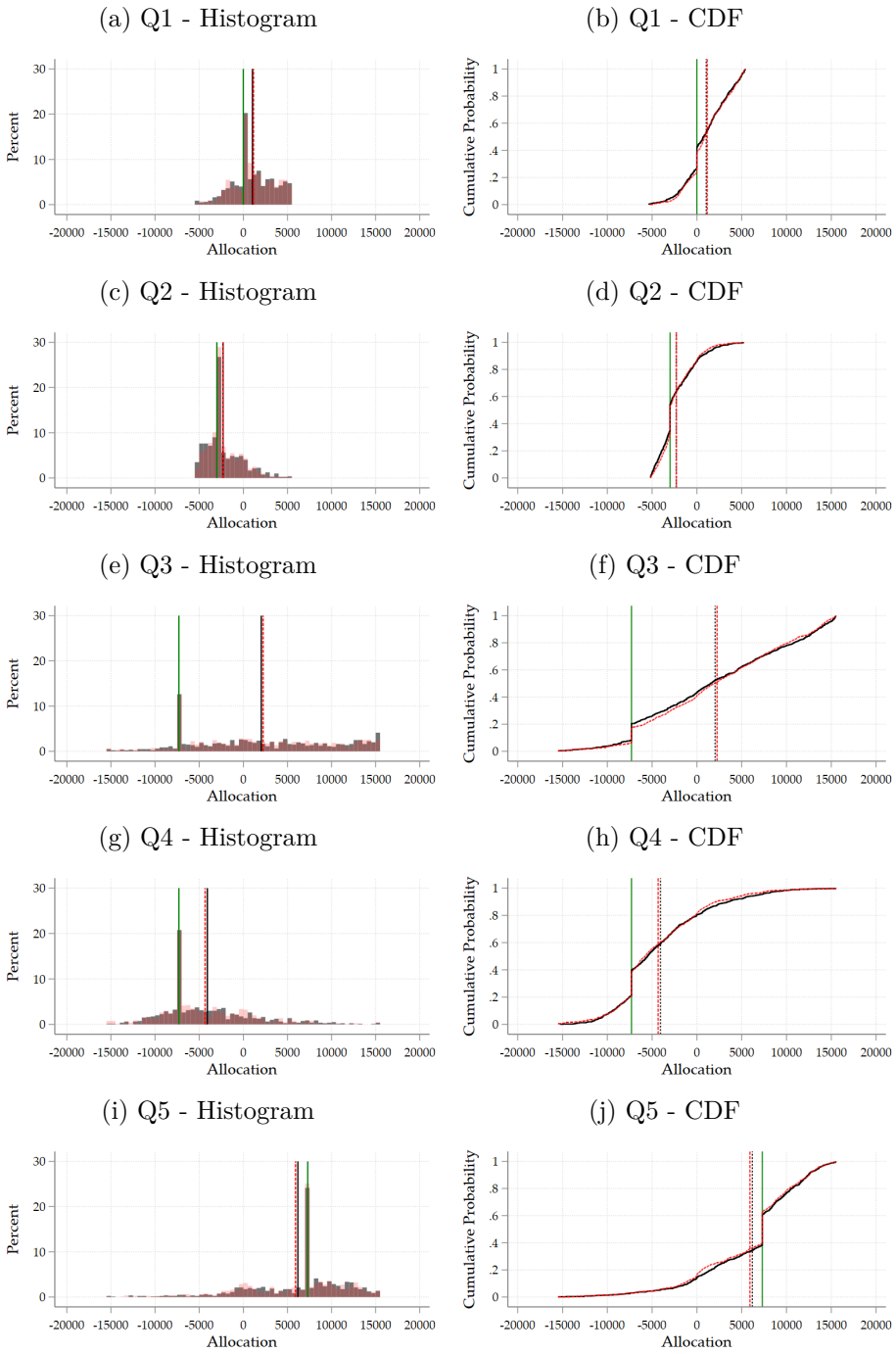
To test for heterogeneous treatment effects, we use equation (2) in split sample analyses for 14 pre-specified demographic sub-groups. For every population subgroup, we test one hypothesis for each of the five vignettes, i.e., we test 70 hypotheses in total. We summarize the resulting information in Table 5 by showing the number of hypotheses that are rejected at the 5% and 10% levels of statistical significance, respectively. As expected, without considering multiple hypothesis testing, some statistically significant differences emerge. For example, when considering model p-values, two out of the 70 tested hypotheses are found to be different from each other at the 5%-level. These differences vanish once we correct for multiple hypothesis testing.[7] This result suggests that the similarity of distributional choices between hypothetical and incentivized scenarios holds across a broad range of population subgroups.

Robustness. We implement a series of robustness checks to analyze the sensitivity of the previous findings.

First, one may be worried that our results are driven by low-quality answers from inattentive respondents. To address this concern, we implemented a control question to screen out

[7]The adjustment for multiple hypothesis testing is made at the level of the population subgroup, i.e., we adjust for the fact that we test five hypotheses within each demographic subgroup.

Figure 3: Distribution of Allocations, "Real-Person Treatment"



Note: This figure displays how distributional choices vary between treatment and control groups. We focus on the first 5 questions that are answered by all respondents. The left panel displays histograms using fixed bins of \$500. The right panel displays cumulative distribution functions. Average choices of the treatment (control) group are represented by dashed red (gray) vertical lines. Solid green vertical lines visualize the status quo distribution of labor market earnings.
Source: Own calculation based on survey responses.

Table 5: Allocations, "Real-Person Treatment", by Demographic Subgroups

Variable	Value	N	Rejected Hypotheses at 5% (10%)		
			Model p-value	Resample p-value	RW p-value
Gender	Male	772	0 (1)	0 (0)	0 (0)
	Female	817	0 (0)	0 (0)	0 (0)
Age	$<$ 45 years	693	0 (0)	0 (0)	0 (0)
	$\geq$ 45 years	909	1 (2)	0 (2)	0 (0)
Ethnicity	White/Caucasian	1299	0 (1)	0 (1)	0 (0)
	Other	299	0 (0)	0 (0)	0 (0)
Education	$\leq$ HS	709	1 (1)	1 (1)	0 (0)
	$>$ HS	893	0 (0)	0 (0)	0 (0)
Employment	(Self)Employed	992	0 (0)	0 (0)	0 (0)
	Other	610	0 (0)	0 (0)	0 (0)
Working Hours	$<$ 35 hours	712	0 (0)	0 (0)	0 (0)
	$\geq$ 35 hours	756	0 (0)	0 (0)	0 (0)
Earnings	$<$ \$40,000	941	0 (0)	0 (0)	0 (0)
	$\geq$ \$40,000	658	0 (0)	0 (0)	0 (0)

Note: This table presents results of the regression analysis outlined in equation (2) for a subsample of respondents based on a particular demographic characteristic. We present the number of respondents, the number of rejected hypotheses according to heteroskedasticity robust model p-values, resample p-values, and p-values adjusted for multiple hypothesis testing. In total, 5 hypotheses are tested for each subsample. Detailed information on regression results are shown in Appendix Tables B8 - B21.
Source: Own calculations based on survey responses.

inattentive respondents at the beginning of the survey. In addition, we pre-specified two alternative sample selection criteria that allow us to filter out low-quality responses. On the one hand, we excluded respondents who did not pass the training question in the preference module on the first try. On the other hand, we excluded respondents with extreme values in the response time distribution, i.e., we dropped respondents above the 90th percentile and below the 10th percentile of the module-specific response time distribution. Table 6 shows that none of the two restrictions alters the results substantially, suggesting that our results are not driven by low-quality responses from inattentive respondents.

Second, one may conjecture that treatment effects become more pronounced with the length of the survey since the potential real-life consequences encourage respondents to stay attentive for a longer time. We can test this conjecture by looking at differences in treatment effects across the length treatment groups (see Figure 1). In particular, we run split sample analyses for these three groups. We summarize the resulting information in Table 7 by showing the number of hypotheses that are rejected at the 5% and 10% levels of statistical significance, respectively. We cannot detect any significant treatment effects once multiple hypothesis testing is accounted for, irrespective of survey length. This result suggests that the answer quality of hypothetical survey modules on fairness preferences does not deteriorate with survey length.

Third, one may suspect that respondents exert more effort on the task if their decision has

Table 6: Allocations, "Real-Person Treatment", Alternative Samples

Panel A: Directly Passed Training Question

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	1154	168.87	1308.25	0.237	0.244	0.686
Q2	1154	31.76	-2574.27	0.774	0.777	0.949
Q3	1154	112.27	3320.52	0.813	0.813	0.949
Q4	1154	-147.78	-4508.78	0.613	0.629	0.935
Q5	1154	-290.95	6884.88	0.361	0.339	0.774

Panel B: Exclude Response Time Outliers

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	924	162.99	1394.95	0.317	0.321	0.766
Q2	924	109.22	-2699.89	0.364	0.375	0.766
Q3	924	-270.11	4030.09	0.606	0.607	0.854
Q4	924	-30.23	-4623.23	0.925	0.933	0.933
Q5	924	-477.60	7159.91	0.169	0.177	0.572

Note: This table presents results of the regression analysis outlined in equation (2) for different restricted samples. Panel A displays results for the sample of respondents that passed the training question of the fairness preference module at the first try. Panel B presents results for a sample that excludes respondents with high (above p90) and low (below p10) response times of the fairness preference module. We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications. See Appendix Tables B6-B7 for the corresponding balancing and Kolmogorov-Smirnov tests.
Source: Own calculation based on survey responses.

Table 7: Allocations, "Real-Person Treatment", by Survey Length Subgroups

		Rejected Hypotheses at 5% (10%)		
Module Length	N	Model p-value	Resample p-value	RW p-value
5 Questions	534	0 (1)	0 (1)	0 (0)
7 Questions	536	0 (0)	0 (0)	0 (0)
9 Questions	532	1 (1)	1 (1)	0 (0)

Note: This table presents results of the regression analysis outlined in equation (2) for a subsample of respondents based on the survey module length, i.e, whether respondents answered five, seven, or nine questions in the preference module. We present the number of respondents, the number of rejected hypotheses according to heteroskedasticity robust model p-values, resample p-values, and p-values adjusted for multiple hypothesis testing. In total, we test five, seven, and nine hypotheses for each subsample depending on the number of questions. Detailed information on regression results are shown in Appendix Tables B22 - B24.
Source: Own calculations based on survey responses.

real-life consequences. Although we do not find any differences in choices in our setting, increased effort could lead to differential choices in settings that are more complicated than ours. To investigate this possibility, we use response times at the question level as a noisy measure for unobserved effort in the task and analyze how response times vary between treatment and control groups. Results in Table 8 show that there are virtually no differences

Table 8: Preferences, Response Time (Min.), "Real-Person Treatment"

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	1585	0.00	1.00	0.892	0.904	0.989
Q2	1585	0.00	0.46	0.936	0.940	0.989
Q3	1585	0.00	0.50	0.837	0.839	0.989
Q4	1585	-0.01	0.37	0.501	0.555	0.936
Q5	1585	0.01	0.36	0.729	0.743	0.981

Note: This table presents results of the regression analysis outlined in equation (2) using the response time in minutes as the dependent variable. For every question, we focus on response times below the 99th percentile of the question-specific response time distribution. We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculation based on survey responses.

between treatment and control groups, suggesting that the effort of respondents does not decrease when facing hypothetical instead of incentivized scenarios.[8]

In this section, we have shown that there are no systematic differences in distributional choices between hypothetical and incentivized scenarios. This conclusion holds for average allocations, distributions of allocations, and within various demographic subgroups. Furthermore, this conclusion is robust to various sensitivity checks, such as excluding low-quality responses. In summary, the results suggest that hypothetical vignettes capture the same fairness preferences as their incentivized analogs.

4 Stability of Fairness Preferences

In this section, we investigate whether the survey module captures genuine fairness preferences that are stable over time. In particular, we use the longitudinal variation between the baseline wave and the follow-up wave of the survey. The follow-up wave consists of a fairness preference module with six questions, three of which are repetitions from the baseline survey. To avoid respondents anchoring their responses on their answers in the baseline survey, we obfuscate the repeated questions by mixing them in random order with novel questions that have not been shown to respondents previously. We present results in three steps. First, we present

[8]To limit the influence of extreme outliers, we focus on respondents whose response time for a particular question is below the 99th percentile of the question-specific response time distribution. In Appendix Table B25, we repeat the exercise without this restriction. Treatment effects increase due to single outliers in the treatment and control groups. However, none of the differences is statistically significant, and our general conclusion remains unaffected.

intertemporal correlations based on the pooled follow-up sample. Second, we investigate whether these intertemporal correlations vary by treatment status in the baseline survey. Third, we present robustness analyses. We registered the follow-up survey in our pre-analysis plan. However, since the survey provider expressed considerable uncertainty about the likely response rates, we did not pre-specify the associated analyses presented in this section.

Intertemporal correlations. We estimate intertemporal correlations through ordinary least-squares using the following model:

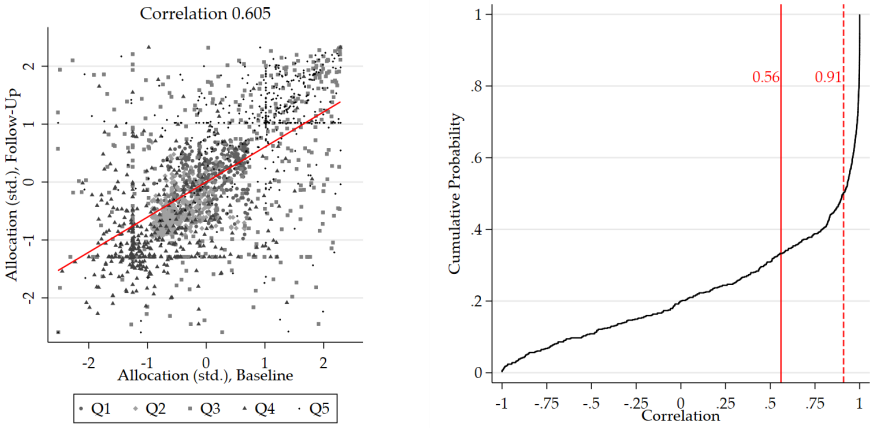
$$\Delta y_{ij,t} = \alpha + \sigma \, \Delta y_{ij,t-1} + \varepsilon_{ij}, \tag{3}$$

where Δy_{ij} is again the difference in allocations to vignette persons A and B by respondent i in vignette j in the baseline wave $(t-1)$ and the follow-up wave (t), respectively. In all estimations, we standardize $\Delta y_{ij,t}$ and $\Delta y_{ij,t-1}$ on the estimation samples such that they have a mean of zero and a standard deviation of one. As a result, estimates of σ can be interpreted as intertemporal correlation coefficients.

Figure 4a plots the raw standardized data of $\Delta y_{ij,t}$ against $\Delta y_{ij,t-1}$, with the fitted line indicating the point estimate of σ. The intertemporal correlation is estimated at 0.61, suggesting sizable stability of distributional choices over time.

Figure 4: Stability of Fairness Preferences, Correlation

(a) Correlation (b) CDF



Note: This figure displays the intertemporal correlation between the baseline and follow-up wave (Figure 4a) and the cumulative distribution function of within-respondent correlations (Figure 4b). In Figure 4a, variables are standardized on the full sample, and the line indicates the line of best fit from a linear regression. In Figure 4b, variables are standardized at the individual level, and the solid (dashed) line indicates the mean (median) correlation across respondents.
Source: Own calculation based on survey responses.

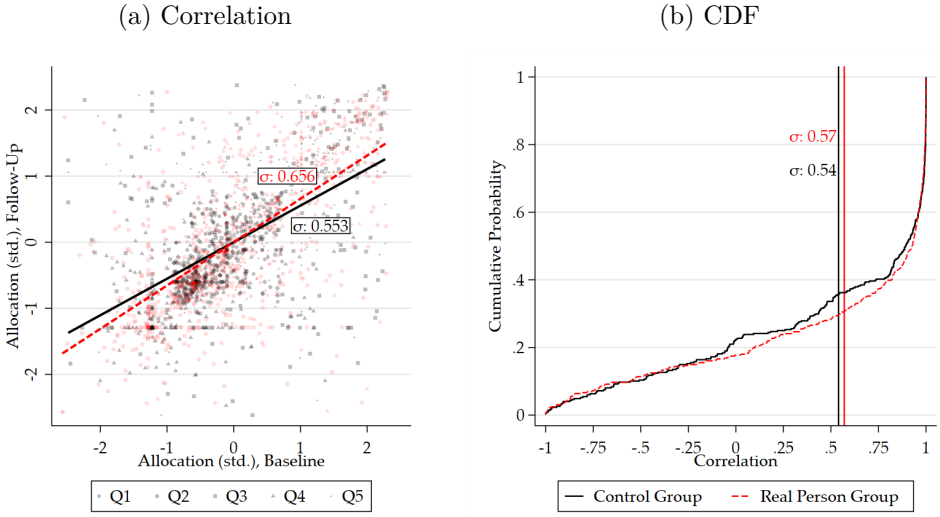
Figure 4b visualizes the cumulative distribution of intertemporal correlations at the individual level. For each respondent, estimates of σ are based on the three repeated questions from baseline and follow-up. More than 80% of respondents display a positive correlation, and

more than 65% have a correlation of 0.50 or higher. The mean (median) correlation across respondents is around 0.56 (0.91). These high intra-respondent correlations reaffirm our conclusion that the recovered distributional choices are fairly stable over time for most respondents in our sample.

Effect of "real-person treatment." The estimated intertemporal correlations in equation (3) may be attenuated by measurement error in $\Delta y_{ij,t-1}$, i.e., the distributional choices in the baseline wave. Therefore, we can use estimates of σ to assess whether incentivized survey questions increase the signal-to-noise ratio in the recovered fairness preferences. If measurement error in $\Delta y_{ij,t-1}$ was less pronounced in incentivized scenarios, σ would be significantly higher in the treatment group than in the control group. Such a finding would suggest that incentivized survey modules yield less noisy estimates of fairness preferences.[9]

In Figure 5a, we replicate Figure 4a by splitting our sample into the treatment and control groups from the baseline wave. Estimates of σ are slightly higher in the treatment (0.66) than in the control group (0.55). The difference of 0.10 is statistically significant at the five percent level (p-value=0.02). In Figure 5b, we show that the difference in stability is less pronounced when considering correlations at the individual level. The average intra-individual correlation is still slightly higher in the treatment (0.57) than in the control group (0.54). However, the difference of 0.03 is not statistically significant at conventional levels of statistical significance (p-value=0.49). These patterns suggest that incentivized survey modules may yield slightly

Figure 5: Stability of Fairness Preferences, "Real-Person Treatment"



Note: This figure displays the intertemporal correlation between the baseline and follow-up wave (Figure 5a) and the cumulative distribution function of within-respondent correlations (Figure 5b) separately for treatment and control groups. In Figure 5a, variables are standardized at the group level, and solid lines indicate lines of best fit from a linear regression. In Figure 5b, variables are standardized at the individual level, and solid lines indicate mean correlations across respondents.
Source: Own calculation based on survey responses.

less noisy estimates of fairness preferences. However, these gains are relatively moderate and

[9] In the follow-up wave, all questions were hypothetical. Since we use $\Delta y_{ij,t}$ as outcomes in equation (2), the associated (classical) measurement error will not bias our estimates of σ.

may be quickly outweighed by the benefits of an unincentivized survey, e.g., lower cost, the potential to target broader population samples, etc.

Robustness. We again implement a series of robustness checks to analyze the sensitivity of the previous findings. These robustness checks are summarized in Table 9. First, we

Table 9: Stability of Fairness Preferences, Correlation, Robustness

	Full Sample	Restricted Sample		
		Training Question	Response Time	Not at Status Quo
Aggregate	0.605 (2127)	0.644 (1710)	0.609 (1707)	0.541 (1206)
Q1	0.375 (424)	0.389 (338)	0.391 (338)	0.413 (295)
Q2	0.328 (426)	0.335 (346)	0.298 (340)	0.436 (199)
Q3	0.407 (425)	0.390 (334)	0.399 (337)	0.356 (326)
Q4	0.332 (427)	0.337 (341)	0.312 (352)	0.317 (193)
Q5	0.382 (425)	0.377 (351)	0.335 (340)	0.497 (193)

Note: This table displays intertemporal correlations between the baseline and follow-up waves for the full sample and three restricted samples. The first restriced sample focuses on respondents that passed the training question of the fairness preference module in the baseline wave at the first try. The second restricted sample excludes respondents with high (above p90) and low (below p10) response times of the fairness preference module in the baseline wave. The third restricted sample excludes respondents whose allocated shares are at most 5 percentage points away from the status quo distribution of earnings. Variables are standardized on the sample used in the corresponding regression. Sample sizes are shown in parenthesis.
Source: Own calculations based on survey responses.

check whether intertemporal correlations change when excluding low-quality answers from inattentive respondents who do not pass the training question on the first try. Intertemporal correlations increase slightly but remain very close to our full sample estimate. In an alternative test, we exclude respondents in the tails of the response time distribution. This sample restriction has virtually no effect on the estimated intertemporal correlations.

Second, we check whether the estimated intertemporal correlations are especially driven by individuals who always leave the slider close to its original position. In particular, we exclude observations where respondents leave the vignette slider within a two-sided five percentage point band around the initial earnings distribution in the baseline and the follow-up. Indeed, there seems a slight drop in intertemporal correlations when excluding these respondents. However, we also emphasize that the implemented test is likely too stringent. On the one hand, we exclude respondents who leave the slider unaltered in bad faith. On the other hand, we also exclude respondents with genuine libertarian preferences. Therefore, we interpret the still substantial intertemporal correlation as a positive signal that we can recover stable preferences in areas further away from initial income positions.

Third, all previous conclusions hold when calculating intertemporal correlations at the level of individual questions. In our previous discussion, we especially focused on intertemporal correlations at the individual level. This is the appropriate level of analysis since factorial survey designs mostly use intra-respondent variation across multiple vignettes to identify the relevant preferences (Wiswall and Zafar, 2018). However, depending on the design, researchers may want to infer preferences from fewer vignettes per individual than in our setting. In

Table 9, we, therefore, assess the extreme case where preferences would be identified based on a single question only. In this case, the preference signal is more noisy, translating into lower intertemporal correlations. Nonetheless, even in the extreme case of using only one vignette, the correlations are still substantial, ranging from 0.33 to 0.41 in the full sample.

In this section, we have shown that the distributional choices are relatively stable over time. This conclusion is robust to various sensitivity checks, among others, excluding low-quality responses. The presence of incentives slightly decreases the noise in elicited preferences. This decrease in noise, however, is fairly moderate and may be quickly outweighed by the potential benefits of running an unincentivized survey. In summary, the results suggest that hypothetical distribution tasks can yield high-quality data on stable fairness preferences.

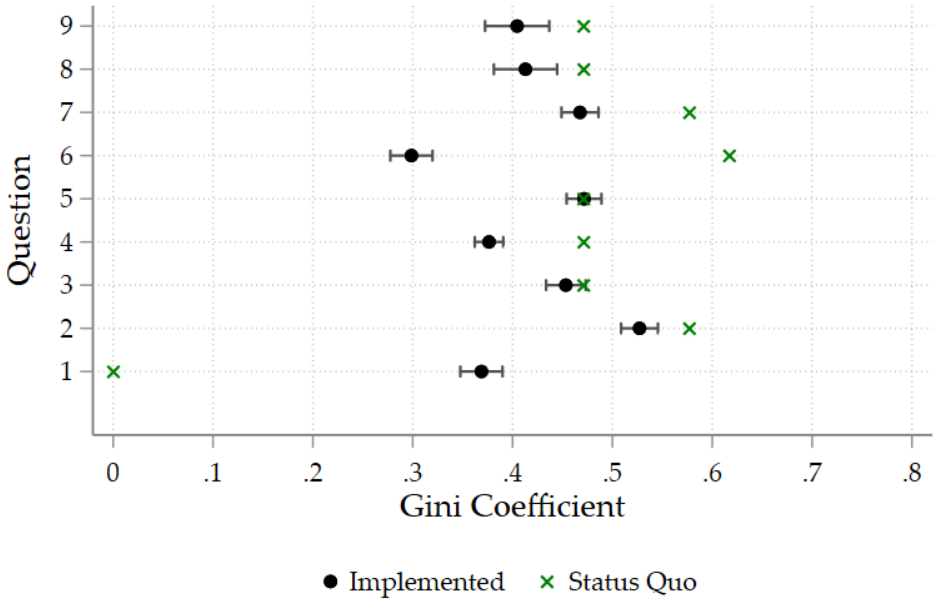
5 Nature of Fairness Preferences

In this section, we accompany the main methodological validation of the previous sections by giving some suggestive insights into the nature of elicited fairness preferences. To be sure, this analysis comes with caveats. The primary purpose of this paper is to assess the measurement of fairness preferences in hypothetical settings as compared to the "gold standard" of incentivized experiments. Therefore, we made several methodological choices that prevent a full substantive analysis of the recovered preferences. For example, to maximize statistical power to detect differences between the treatment and control group, we show all respondents the same randomly selected subset of vignettes. Consequently, vignette characteristics are not equally represented, and correlations may exist among them. These features may affect respondents' willingness to tolerate inequality and how they incorporate different vignette characteristics into their choices. Therefore, we view the following analysis as a suggestive test of whether the recovered preferences are broadly consistent with findings from the existing literature.

With these caveats in mind, we will present the results of this section as follows. First, we analyze the level of inequality implemented by respondents. Second, we will analyze the prevalence of different fairness types in our sample. Lastly, we show the sensitivity of fairness preferences to different characteristics of the evaluated vignette persons. In all analyses, we will focus on unincentivized scenarios from the control group. However, our conclusions remain unaffected when focusing on the incentivized sample—see Appendix Figures C4, C6 and Appendix Table C26 for replications of the main exhibits of this section based on the treatment group. The analyses of this section are exploratory. Therefore, they have not been registered in our pre-analysis plan.

Implemented inequality. Figure 6 compares implemented inequality by respondents to the initial inequality separately for each vignette. The implemented Gini coefficients show

Figure 6: Gini Coefficient



Note: This figure compares implemented Gini coefficients (gray dots) to initial Gini coefficients (green crosses) in each vignette. Gray bars indicate 95 percent confidence intervals. Gini coefficients are calculated at the respondent level as $\frac{|x-y|}{x+y}$ where x (y) is the amount allocated to vignette person A (B). We focus exclusively on respondents in the control group, i.e., those respondents who faced hypothetical scenarios. Appendix Figure C4 replicates the analysis for respondents in the treatment group.
Source: Own calculations based on survey responses.

substantial variation across vignettes (0.30–0.53). For comparison, Almås et al. (2020) use a representative sample of American respondents to show that they would implement a Gini coefficient of 0.35 (0.54) if the income-generating process were purely based on luck (merit). This suggests the range of implemented inequality across different scenarios in our setting is plausible.

In Appendix Figure C5, we furthermore illustrate how inequality acceptance varies across respondents with different socio-demographic characteristics. Respondents who are more inequality-accepting tend to be older, more educated, and work longer hours. Those with a lower inequality tolerance tend to be female and non-white. Again, these patterns are broadly consistent with existing literature. For instance, the findings of Almås et al. (2020) indicate that women and individuals with lower educational attainment are less accepting of inequality compared to men and those with higher education. The authors interpret these patterns in the light of potential self-serving biases in fairness preferences. It is reassuring that our survey replicates these patterns as well.

Fairness types. Experimental literature has focused on estimating the prevalence of different fairness types that can be mapped to fairness principles in the philosophical literature—see Almås et al. (2024b) for a recent overview. On one end of the spectrum is the *egalitarian position*. Egalitarians consider all inequalities unfair, regardless of how these inequalities

come about. Therefore, the egalitarian position prescribes an equal income distribution in any distributive situation. At the opposite end of the spectrum is the *libertarian position*. Libertarians consider all inequalities fair regardless of how these inequalities come about (Nozick, 1974). Therefore, the libertarian position prescribes a distribution of income that corresponds to the initial distribution in any distributive situation. Between these two extreme positions, there are several *intermediate positions*, such as the responsibility-sensitive positions proposed by Arneson (1989), Cohen (1989), Dworkin (1981a), and Dworkin (1981b). These intermediate positions advocate for distinguishing between different sources of inequality, such as discretionary choices, ability, preferences, or circumstantial factors.

We estimate the prevalence of the egalitarian position by calculating the share of respondents who implement equal splits in all vignettes. Similarly, we estimate the prevalence of the libertarian position by calculating the share of respondents who accept initial inequality in all vignettes. When calculating these shares, we allow for "trembling hand" mistakes (Choi et al., 2007). For our baseline estimates, we use two-sided five percentage point bands around the egalitarian and libertarian answers to a vignette and allow respondents to be outside of the corresponding band for at most one vignette without repercussions on their classification as egalitarians or libertarians. We estimate the prevalence of the intermediate position as the remaining share of respondents who are not classified as egalitarians or libertarians.

Table 10 shows the results, where the highlighted areas represent our baseline estimates. Around two percent of respondents are classified as egalitarian, whereas around nine percent are classified as libertarians. The remaining 89% percent of respondents adopt intermediate positions. Therefore, most respondents adopt fairness positions that vary with the characteristics of the respective vignette. We note that this conclusion does not vary with the leniency with which we accept "trembling hand" mistakes. Even in the most lenient specifications where we allow for two-sided ten percentage point bands and two inconsistent answers, the share of respondents adopting intermediate positions is still 70%. Furthermore, we note that conditional on the adopted rule for "trembling hand" mistakes, the presented estimates for the prevalence of egalitarian and libertarian positions should be interpreted as upper bounds. We only presented respondents with a limited selection of five to nine vignettes. Therefore, in additional questions, the number of divergences from the egalitarian and libertarian positions can stay constant at best but not decrease. The estimated shares of egalitarians and libertarians in the US are smaller than the corresponding shares estimated in Almås et al. (2020). Their estimates classify 15% and 29% of the US population as egalitarians and libertarians, respectively. This difference may be rationalized by variations in how different preference types are identified or by the increased richness of the distributional scenarios in our setting.[10] Since the vignettes provide multidimensional information on the earnings-relevant characteristics of the recipients, respondents can express positions that deviate from the polar cases of egalitarian/libertarian fairness preferences in more nuanced

[10] Almås et al. (2020) identify egalitarians as respondents who distribute resources equally in a situation where initial inequality is driven by productivity. They identify libertarians as respondents who do not redistribute at all when initial inequality is purely based on luck.

ways.

Table 10: Preference Types

	Share Egalitarians (%)			Share Libertarians (%)		
Max. abs. difference (pp)	2	5	10	2	5	10
Allow for 0 inconsistent answers	0.00	0.51	1.73	5.68	6.41	7.65
Allow for 1 inconsistent answers	0.00	**1.46**	3.75	7.89	**9.47**	13.84
Allow for 2 inconsistent answers	0.89	2.71	7.69	10.75	13.90	22.22

Note: This table presents shares of egalitarians (libertarians) according to consistent choices in all questions of the baseline wave. We also vary the leniency of the classification by allowing for 0, 1, 2 answers that are inconsistent with egalitarian (libertarian) choices. In the baseline (highlighted estimates), we allow for a deviation of +/-5 percentage points and inconsistent choices in one question only. We focus exclusively on respondents in the control group, i.e., those respondents who faced hypothetical scenarios. Appendix Table C26 replicates the analysis for respondents in the treatment group.
Source: Own calculations based on survey responses.

Importance of vignette person characteristics. In the last step, we analyze the impact of particular earnings-relevant characteristics on the fairness preferences of respondents. To this end, we transform our data as follows. We create a data set where each row represents one person m from vignette j. Then, we replicate these data for each respondent i who made a distributional choice for vignette j and include the corresponding income allocations $y_{im(j)}$ as the outcome variable of interest. Stacking these data, we obtain a panel data set with multiple observations for each vignette person $m(j)$ and each respondent i.

We then estimate the following model via ordinary least-squares:

$$\begin{aligned} \ln y_{im(j)} = {} & \beta_1\, gender_{m(j)} + \beta_2\, age_{m(j)} + \beta_3\, educ_{m(j)} \\ & + \beta_4\, educpar_{m(j)} + \beta_5\, hours_{m(j)} + \beta_6\, \ln earn_{m(j)} \\ & + \theta_{[earn_{A(j)}+earn_{B(j)}]} + \epsilon_{im(j)}. \end{aligned} \tag{4}$$

The right-hand side variables in the first two lines of equation (4) represent the six vignette characteristics considered in our fairness preference module. The associated coefficients $\beta_1 - \beta_6$ capture the linear effect of personal characteristics on fair earnings conditional on the remaining vignette characteristics. We control non-parametrically for the total sum of vignette earnings $earn_{A(j)} + earn_{B(j)}$ by including corresponding fixed effects. Thereby, we account for the fact that vignette persons who are paired with high-earning persons in their vignette may mechanically receive higher income allocations. Standard errors are clustered at the respondent level.

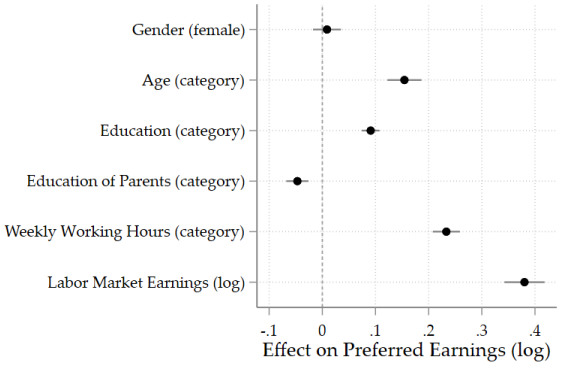
The interpretation of the estimated coefficients comes with two important caveats. First, we cannot control how respondents associate the displayed vignette characteristics with other unobserved characteristics, such as productivity or job performance. Therefore, $\beta_1 - \beta_6$ are composite parameters that capture the reward for a certain characteristic, and the reward

for unobserved factors correlated with it while holding all other observed characteristics constant. Second, we disregard non-linear effects across the ordinal vignette characteristics. The methodological choices mentioned at the beginning of this section limit the available variation in our data and prevent us from relaxing this stringent functional form assumption.

Figure 7a displays the coefficients from regression equation (4) along with the corresponding 95% confidence intervals. On the one hand, the results indicate that fair labor market earnings

Figure 7: Fairness Preferences, Importance of Profile Characteristics

(a) Regression Analysis



(b) Text Analysis

Note: This figure displays how respondents take the vignette characteristics into account for their distributional choices. Figure 7a displays the point estimates and 95 percent confidence intervals from equation (4). Figure 7b displays a word cloud from a text analysis of the open-ended question in the fairness preference module. At the end of the module, people were asked about how they came up with their distributional choices and to describe their reasoning in their own words. Based on the text corpus from the open-ended answers, we used natural language processing techniques to rank the frequency of specific terms. We focus exclusively on respondents in the control group, i.e., those respondents who faced hypothetical scenarios. Appendix Table C27 replicates the analysis for different transformations of the outcome variable, and under the inclusion of respondent fixed effects. Appendix Figure C6 replicates the analysis for respondents in the treatment group.
Source: Own calculation based on survey responses.

increase with vignette characteristics relevant to labor market performance and that are (partially) under the control of individuals. For example, there are strong positive effects of education and weekly working hours on fair income allocations. This result suggests that fairness preferences in the US are at least partially consistent with normative theories that emphasize the role of discretionary choices in determining fair income shares, (e.g., Konow, 2000). Similar conclusions can be drawn for age and initial earnings, if interpreted as proxy indicators for relevant labor market experience and on-the-job performance, respectively.[11]

On the other hand, the results show that non-discretionary vignette characteristics are not rewarded in fair income allocations. For example, conditional on the other vignette

[11] By controlling for initial labor market earnings, we ensure that we measure the desired reward of all other characteristics, holding initial labor market earnings constant. For instance, a positive coefficient on education implies that conditional on initial market earnings, higher-educated people are considered more deserving than lower-educated people. In a full roll-out of our factorial survey, initial labor market earnings would be uncorrelated with all other vignette characteristics and we could replicate our analyses excluding initial labor market earnings. However, due to the methodological focus of our data collection, we operate with a limited number of vignettes where the characteristics of interest are correlated with each other. To account for such correlations, we include initial labor market earnings in all our estimations.

characteristics, the point estimate of gender cannot be distinguished from zero, suggesting that US residents perceive adjusted gender gaps in labor market earnings as unfair. Furthermore, the respondents in our sample assign higher (lower) earnings to individuals with lower (higher) parental education. This finding could be rationalized by the fact that respondents are willing to compensate people for a disadvantaged socio-economic family background. Note, however, that our analysis only presents relative effects of vignette characteristics. Therefore, we cannot distinguish whether respondents compensate individuals for a disadvantaged background (low parental education) or penalize individuals coming from an advantaged family background (high parental education).

Given the abovementioned caveats, we interpret these findings as suggestive. However, in Appendix Table C27, we show that all previous conclusions are robust to alternative specifications. In particular, we use preferred income shares or preferred income ratios as outcome variables of interest. While the magnitude of coefficients changes, the direction of effects remains unaltered. Also, our results do not change qualitatively when controlling for respondent fixed effects, i.e., when we only use within-respondent variation to estimate preferred rewards. In addition, while our baseline analysis only uses responses from respondents with hypothetical tasks, the measured rewards are similar in the real-person treatment group (Appendix Figure C6).

To substantiate the quantitative evidence, we also use natural language processing techniques to present findings from a text analysis (Ferrario and Stantcheva, 2022). At the end of the baseline survey, we asked respondents how they made their distributional choices and allowed them to describe their reasoning in an open-text field. Figure 7b visualizes the text analysis in a word cloud, highlighting the frequency of observed terms. The word cloud shows that respondents put a strong emphasis on working hours, earnings, and the education of the vignette persons when making their distributional choices. The emphasis on these characteristics, therefore, echoes the results from our quantitative analysis.

In this section, we have shown that our survey module recovers fairness preferences that are broadly consistent with the existing literature. This conclusion holds for the degree of inequality acceptance, the prevalence of fairness types, and the characteristics determining the extent of fair income allocations. Since the data collection was designed for methodological validation, we urge readers to treat these substantive results cautiously. However, the results point to the ability of our survey module to uncover nuanced fairness positions and to describe fairness preferences in societies more broadly.

6 Conclusion

This study validates a novel survey tool designed to measure fairness preferences using realistic yet hypothetical scenarios.

We conduct this validation using a two-wave survey covering a representative sample of the US population. Our results demonstrate that fairness preferences are not influenced by the prospect of real-world implementation, even when monetary stakes are high. This conclusion holds true for both the general population and across various demographic subgroups. Moreover, comparing individual responses across the two waves reveals that fairness preferences are stable over time, regardless of whether they originate from hypothetical or incentivized scenarios. We furthermore provide suggestive evidence that the elicited preferences are consistent with established findings on fairness preferences in the US.

Therefore, our validation provides compelling evidence that fairness preferences from hypothetical surveys are not "just cheap talk." Instead, they can yield credible insights into the nature and anatomy of these preferences. We emphasize that these conclusions are context-dependent. Therefore, we currently plan additional validation exercises to show that our findings extend beyond the context of *WEIRD* countries like the US.[12]

[12]Henrich et al. (2010) show that behavior in experiments varies substantially across cultural contexts and that evidence from Western, Educated, Industrial, Rich, and Democratic (WEIRD) countries can hardly be extrapolated to other settings.

References

Adriaans, J. (2023). "Fairness of Earnings in Europe: The Consequences of Unfair under- and Overreward for Life Satisfaction." *European Sociological Review* 39 (1), pp. 118–131.

Alesina, A., S. Stantcheva, and E. Teso (2018). "Intergenerational Mobility and Preferences for Redistribution." *American Economic Review* 108 (2), pp. 521–554.

Almås, I., A. W. Cappelen, E. Ø. Sørensen, and B. Tungodden (2024a). *Fairness Across the World: Preferences and Beliefs*. Mimeo.

Almås, I., A. W. Cappelen, and B. Tungodden (2020). "Cutthroat Capitalism versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians?" *Journal of Political Economy* 128 (5), pp. 1753–1788.

Almås, I., P. Hufe, and D. Weishaar (2023). "Equality of Opportunity: Fairness Preferences and Beliefs About Inequality." *Handbook of Equality of Opportunity*. Ed. by M. Sardoč. Cham: Springer International Publishing, pp. 1–23.

– (2024b). *Experimental Evidence on Attitudes Toward Inequality and Fairness*. CEPR Discussion Paper 19620.

Andersen, A. G., S. Franklin, T. Getahun, A. Kotsadam, V. Somville, and E. Villanger (2023). "Does Wealth Reduce Support for Redistribution? Evidence from an Ethiopian Housing Lottery." *Journal of Public Economics* 224, p. 104939.

Andre, P. (forthcoming). "Shallow Meritocracy." *Review of Economic Studies*.

Arneson, R. J. (1989). "Equality and Equal Opportunity for Welfare." *Philosophical Studies* 56 (1), pp. 77–93.

Auspurg, K. and T. Hinz (2014). *Factorial Survey Experiments*. Thousand Oaks: SAGE Publications.

Auspurg, K., T. Hinz, S. Liebig, and C. Sauer (2015). "The Factorial Survey as a Method for Measuring Sensitive Issues." *Improving Survey Methods: Lessons from Recent Research*. New York: Routledge/Taylor & Francis Group, pp. 137–149.

Auspurg, K., T. Hinz, and C. Sauer (2017). "Why Should Women Get Less? Evidence on the Gender Pay Gap from Multifactorial Survey Experiments." *American Sociological Review* 82 (1), pp. 179–210.

Bauer, M., J. Chytilová, and E. Miguel (2020). "Using Survey Questions to Measure Preferences: Lessons from an Experimental Validation in Kenya." *European Economic Review* 127, pp. 1–8.

Bick, A., A. Blandin, and R. Rogerson (2022). "Hours and Wages." *The Quarterly Journal of Economics* 137 (3), pp. 1901–1962.

Blau, F. D. and L. M. Kahn (2017). "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature* 55 (3), pp. 789–865.

Cappelen, A. W., A. D. Hole, E. Ø. Sørensen, and B. Tungodden (2007). "The Pluralism of Fairness Ideals: An Experimental Approach." *American Economic Review* 97 (3), pp. 818–827.

Cappelen, A. W., J. Konow, E. Ø. Sørensen, and B. Tungodden (2013). "Just Luck: An Experimental Study of Risk-Taking and Fairness." *American Economic Review* 103 (4), pp. 1398–1413.

Cappelen, A. W., E. Ø. Sørensen, and B. Tungodden (2010). "Responsibility for What? Fairness and Individual Responsibility." *European Economic Review* 54 (3), pp. 429–441.

Choi, S., R. Fisman, D. Gale, and S. Kariv (2007). "Consistency and Heterogeneity of Individual Behavior under Uncertainty." *American Economic Review* 97 (5), pp. 1921–1938.

Cohen, G. A. (1989). "On the Currency of Egalitarian Justice." *Ethics* 99 (4), pp. 906–944.

Costa-Font, J. and F. Cowell (2014). "Social Identity and Redistributive Preferences: A Survey." *Journal of Economic Surveys* 29 (2), pp. 357–374.

Day, B., I. J. Bateman, R. T. Carson, D. Dupont, J. J. Louviere, S. Morimoto, R. Scarpa, and P. Wang (2012). "Ordering Effects and Choice Set Awareness in Repeat-Response Stated Preference Studies." *Journal of Environmental Economics and Management* 63 (1), pp. 73–91.

Dworkin, R. (1981a). "What Is Equality? Part 1: Equality of Welfare." *Philosophy & Public Affairs* 10 (3), pp. 185–246.

– (1981b). "What Is Equality? Part 2: Equality of Resources." *Philosophy & Public Affairs* 10 (4), pp. 283–345.

Enke, B., R. Rodríguez-Padilla, and F. Zimmermann (2022). "Moral Universalism: Measurement and Economic Relevance." *Management Science* 68 (5), pp. 3590–3603.

Falk, A., A. Becker, T. Dohmen, D. Huffman, and U. Sunde (2023). "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences." *Management Science* 69 (4), pp. 1935–1950.

Fallucchi, F., D. Nosenzo, and E. Reuben (2020). "Measuring Preferences for Competition with Experimentally-Validated Survey Questions." *Journal of Economic Behavior & Organization* 178, pp. 402–423.

Fehr, E., T. Epper, and J. Senn (2024). "Social Preferences and Redistributive Politics." *The Review of Economics and Statistics*, pp. 1–45.

Ferrario, B. and S. Stantcheva (2022). "Eliciting People's First-Order Concerns: Text Analysis of Open-Ended Survey Questions." *AEA Papers and Proceedings* 112, pp. 163–169.

Fisman, R., K. Gladstone, I. Kuziemko, and S. Naidu (2020). "Do Americans Want to Tax Wealth? Evidence from Online Surveys." *Journal of Public Economics* 188, p. 104207.

Fisman, R., P. Jakiela, S. Kariv, and S. Vannutelli (2023). "The Distributional Preferences of Americans, 2013–2016." *Experimental Economics* 26 (4), pp. 727–748.

Gaertner, W. and E. Schokkaert (2012). *Empirical Social Choice: Questionnaire-Experimental Studies on Distributive Justice*. Cambridge: Cambridge University Press.

Gaertner, W. and L. Schwettmann (2007). "Equity, Responsibility and the Cultural Dimension." *Economica* 74 (296), pp. 627–649.

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), pp. 1826–1863.

Goldin, C. (2014). "A Grand Gender Convergence: Its Last Chapter." *American Economic Review* 104 (4), pp. 1091–1119.

Haaland, I., C. Roth, and J. Wohlfart (2023). "Designing Information Provision Experiments." *Journal of Economic Literature* 61 (1), pp. 3–40.

Harmon, C., H. Oosterbeek, and I. Walker (2003). "The Returns to Education: Microeconomics." *Journal of Economic Surveys* 17 (2), pp. 115–156.

Henrich, J., S. J. Heine, and A. Norenzayan (2010). "The Weirdest People in the World?" *Behavioral and Brain Sciences* 33 (2-3), pp. 61–83.

Jasso, G. and M. Webster (1997). "Double Standards in Just Earnings for Male and Female Workers." *Social Psychology Quarterly* 60 (1), pp. 66–78.

– (1999). "Assessing the Gender Gap in Just Earnings and Its Underlying Mechanisms." *Social Psychology Quarterly* 62 (4), pp. 367–380.

Konow, J. (1996). "A Positive Theory of Economic Fairness." *Journal of Economic Behavior & Organization* 31 (1), pp. 13–35.

Konow, J. (2000). "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions." *American Economic Review* 90 (4), pp. 1072–1092.

Konow, J., T. Saijo, and K. Akai (2020). "Equity versus Equality: Spectators, Stakeholders and Groups." *Journal of Economic Psychology* 77, p. 102171.

Kuhn, P. and F. Lozano (2008). "The Expanding Workweek? Understanding Trends in Long Work Hours among U.S. Men, 1979–2006." *Journal of Labor Economics* 26 (2), pp. 311–343.

Lemieux, T. (2006). "Postsecondary Education and Increasing Wage Inequality." *American Economic Review* 96 (2), pp. 195–199.

Magnac, T. and S. Roux (2021). "Heterogeneity and Wage Inequalities over the Life Cycle." *European Economic Review* 134, p. 103715.

Mazumder, B. (2005). "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data." *Review of Economics and Statistics* 87 (2), pp. 235–255.

Mollerstrom, J., B.-A. Reme, and E. Ø. Sørensen (2015). "Luck, Choice and Responsibility: An Experimental Study of Fairness Views." *Journal of Public Economics* 131, pp. 33–40.

Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.

Roemer, J. E. and A. Trannoy (2016). "Equality of Opportunity: Theory and Measurement." *Journal of Economic Literature* 54 (4), pp. 1288–1332.

Romano, J. P. and M. Wolf (2005). "Stepwise Multiple Testing as Formalized Data Snooping." *Econometrica* 73 (4), pp. 1237–1282.

– (2016). "Efficient Computation of Adjusted p -Values for Resampling-Based Stepdown Multiple Testing." *Statistics & Probability Letters* 113, pp. 38–40.

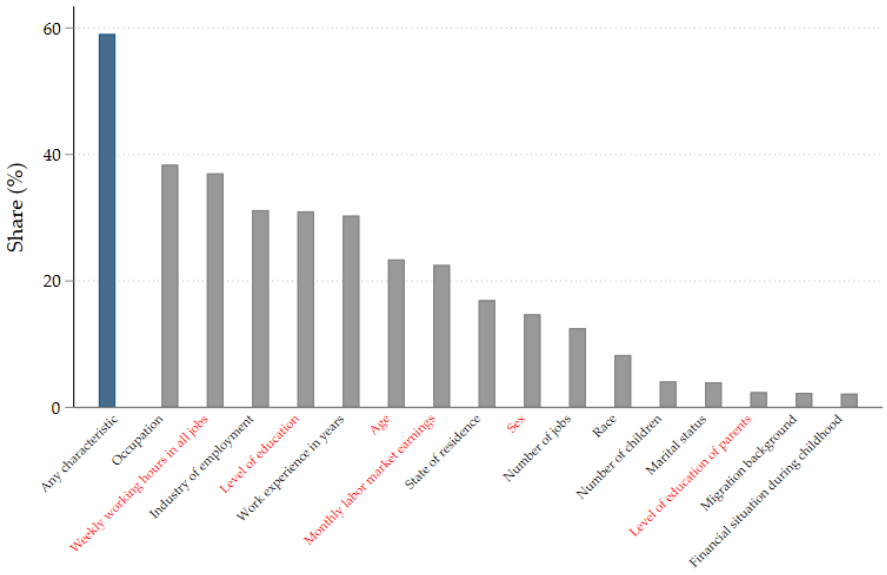
Stantcheva, S. (2023). "How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible." *Annual Review of Economics* 15, pp. 205–35.

Starmans, C., M. Sheskin, and P. Bloom (2017). "Why People Prefer Unequal Societies." *Nature Human Behaviour* 1 (4), pp. 1–7.

Wiswall, M. and B. Zafar (2018). "Preference for the Workplace, Investment in Human Capital, and Gender." *The Quarterly Journal of Economics* 133 (1), pp. 457–507.

A Survey Tool and Data Collection

Figure A1: Distribution Task, Chosen Characteristics



Note: The figure shows results from the fourth module of the baseline survey that refers to other questions about the labor market. The figure illustrates what share of respondents chose to get to know a certain characteristic when asked to decide on the fair earnings split between two persons. In the first step, respondents were informed about the distribution task and asked whether they would like to know anything about the two persons. In the second step, respondents could choose up to five characteristics that they would see for the two persons before redistributing earnings. Characteristics that we chose ex-ante to be included in the main survey module on fairness preferences are highlighted in red.
Source: Own calculations based on survey responses.

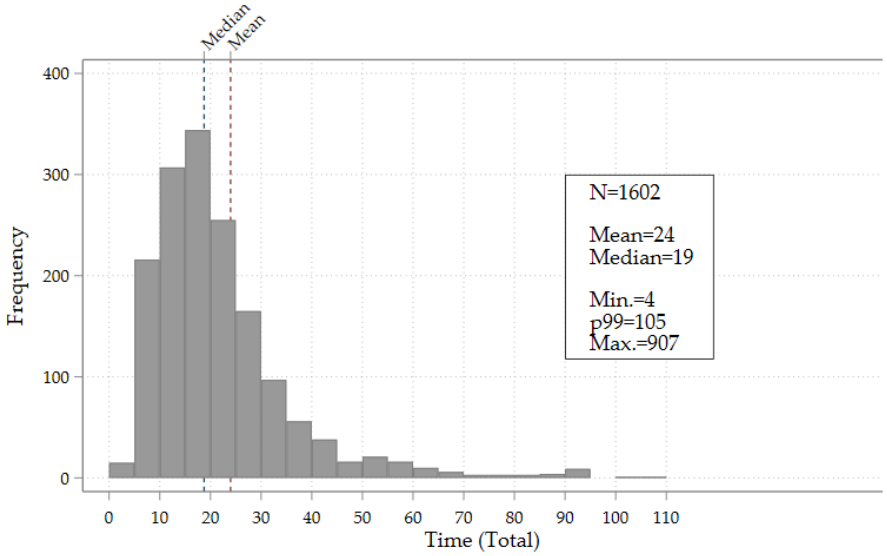
Table A1: Characteristics of the Real Person

Characteristic	Expression
Gender	Female
Age	26
Education	University
Education of Parents	High School
Working Hours	20
Monthly Earnings	\$4,100

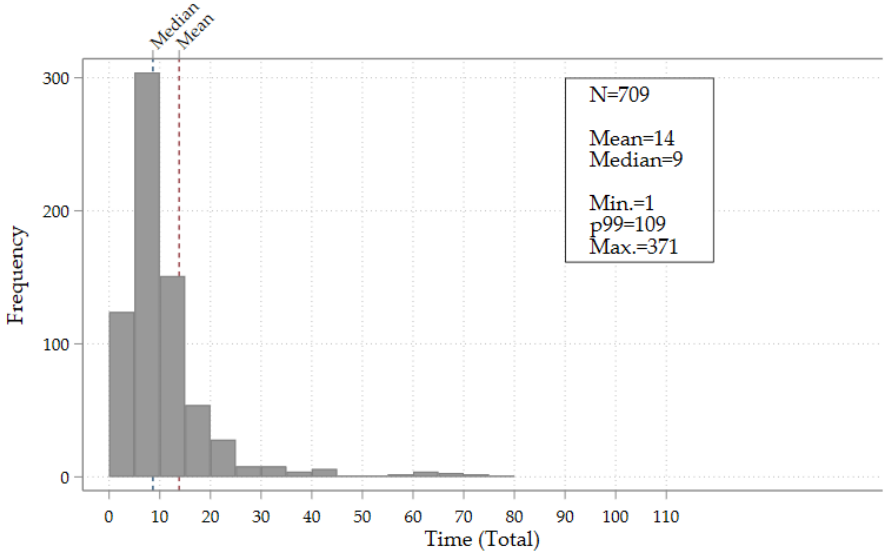
Note: This table displays the expressions of characteristics for the real person who was hired for one month. The initial monthly earnings were drawn randomly from the five available values displayed in Table 1.

Figure A2: Survey Completion Time

(a) Baseline Wave



(b) Follow-Up Wave



Note: This figure displays the frequency distribution of total survey completion time using fixed five-minute bins for the baseline wave (Figure A2a) and the follow-up wave (Figure A2b). For better visibility, the figure only shows the distribution for values below the 99th percentile of the respective survey completion time distribution. The vertical blue (red) lines display the median (mean) of the survey completion time. Further information on the respective distribution is shown in the box next to the histogram.
Source: Own calculations based on survey responses.

Table A2: Demographics

Variable	Share Baseline (%)	Share Follow-Up (%)
Gender		
Male	48.19	47.95
Female	51.00	51.90
Non-Binary	0.81	0.14
Age		
18 - 24	12.98	5.92
25 - 34	13.98	7.33
35 - 44	16.29	13.68
45 - 54	15.61	10.58
55 - 64	18.04	25.25
Older than 64	23.10	37.24
Education		
Less than Middle School	0.37	0.28
Middle School	3.93	3.10
High School Graduate	39.95	35.68
Completed Some College	16.79	14.95
College Degree	25.03	29.06
Master's Degree	10.11	11.42
Doctoral Degree / Law or Professional Degree	3.81	5.50
Ethnicity		
White / Caucasian	81.09	88.15
Black / African American	9.43	5.92
American Indian	1.12	0.42
Alaska Native	0.19	0.00
Asian / Asian American	3.68	3.24
Native Hawaiian	0.31	0.14
Pacific Islander	0.06	0.00
Other	3.87	2.12
Prefer not to answer	0.25	0.00
Hispanic		
Yes	10.05	6.49
No	89.83	93.51
Prefer not to answer	0.12	0.00
Region		
North-East	19.35	24.68
Mid-West	33.52	32.44
West	23.91	18.62
South	23.22	24.26

Continued on next page.

Table A2: Demographics (cont.)

Variable	Share Baseline (%)	Share Follow-Up (%)
Employment Status		
Employed	53.00	47.53
Self-Employed	8.93	8.32
Unemployed	5.74	3.67
Student	3.68	1.13
Retiree	28.65	39.35
Working Hours		
9 hours/week or less	27.53	34.27
10-34 hours/week	16.92	12.41
35-44 hours/week	35.83	35.26
More than 44 hours/week	11.36	9.73
Prefer not to answer	8.36	8.32
Income		
Did not work	22.28	27.79
$0 - $25,000	19.41	12.98
$25,000 - $40,000	17.04	12.55
$40,000 - $59,000	14.17	13.82
$59,000 - $88,000	12.23	14.25
More than $88,000	14.67	18.62
Prefer not to answer	0.19	0.00
Education of Mother		
Less than Middle School	2.00	2.82
Middle School	7.99	8.46
High School Graduate	47.50	48.52
Completed Some College	12.80	8.74
College Degree	19.66	21.02
Master's Degree	7.30	7.19
Doctoral Degree / Law or Professional Degree	1.81	2.26
Prefer not to answer	0.94	0.99
Education of Father		
Less than Middle School	3.43	4.37
Middle School	11.61	14.39
High School Graduate	44.19	42.03
Completed Some College	11.74	9.31
College Degree	17.79	18.90
Master's Degree	5.93	6.21
Doctoral Degree / Law or Professional Degree	3.00	3.24
Prefer not to answer	2.31	1.55

Note: This table presents summary statistics for demographics in the baseline wave (N=1602) and the follow-up wave (N=709). Information on respondent demographics are obtained from the demographics module of the baseline wave of the survey. We also include answers of those respondents that did prefer not to answer specific demographic questions.
Source: Own calculations based on survey responses.

B Effects of the "Real-Person Treatment"

Table B3: Joint Balancing Tests

Demographics	Binary Split	Dep. Variable	N	Point Estimate	Control Mean	RW p-values
Gender	Male vs. Female	TreatB	1589	0.027	0.5006	0.964
		TreatC	1589	-0.007	0.5176	1.000
Age	$<$ 45 vs. $\geq$ 45 years	TreatB	1602	-0.010	0.5725	1.000
		TreatC	1602	-0.050	0.5925	0.418
Ethnicity	White/Caucasian vs. Other	TreatB	1598	0.008	0.1830	1.000
		TreatC	1598	0.006	0.1842	1.000
Education	$\leq$ HS vs. $>$ HS	TreatB	1602	0.005	0.5550	1.000
		TreatC	1602	0.037	0.5387	0.818
Employment	(Self)Employed vs. Other	TreatB	1602	-0.031	0.3962	0.922
		TreatC	1602	-0.018	0.3900	0.996
Hours	$<$ 35 vs. $\geq$ 35 hours	TreatB	1468	-0.003	0.5163	1.000
		TreatC	1468	-0.019	0.5249	0.996
Income	$<$ \$40000 vs. $\geq$ \$40000	TreatB	1599	0.003	0.4098	1.000
		TreatC	1599	-0.003	0.4128	1.000

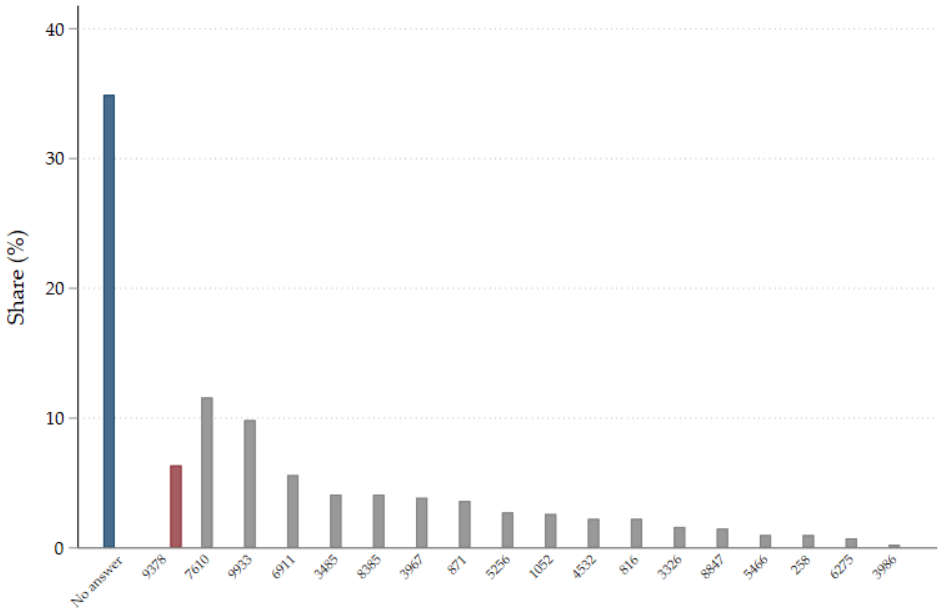
Note: This table presents results of the joint balancing test outlined in the pre-analysis plan. The table shows point estimates for the coefficient of interest, the mean of the control group (first group in binary split), and associated heteroskedasticity robust p-values adjusted for multiple hypothesis testing using the Romano and Wolf (2016) step-down procedure with 1000 bootstrap draws. *TreatB* refers to the treatment in the module on inequality perceptions. *TreatC* refers to the treatment in the module on fairness preferences. Sample sizes differ across demographic characteristics due to the exclusion of responses such as "Prefer not to answer" (see Table 2).
Source: Own calculation based on survey responses.

Table B4: Allocated Shares, "Real-Person Treatment"

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	1602	1.06	59.611	0.324	0.322	0.841
Q2	1602	0.08	27.904	0.931	0.932	0.932
Q3	1602	0.68	56.569	0.591	0.598	0.841
Q4	1602	-0.83	36.864	0.321	0.325	0.841
Q5	1602	-0.85	69.950	0.343	0.337	0.841

Note: This table presents results of the regression analysis outlined in equation (2) where the dependent variable is the share allocated to person A. We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculation based on survey responses.

Figure B3: Guess of the Real Person ID, Shares



Note: This figure displays the shares of different guesses for the real person for those respondents in the preference treatment group with financial incentives ($N = 802$). The blue bar indicates the share of respondents who did not provide any answer, while the red bar represents the share of respondents who guessed the real person correctly. Other guesses are displayed as gray bars. A correct guess was incentivized with $5.
Source: Own calculations based on survey responses

Table B5: Distribution of Allocations, Kolmogorov–Smirnov Test

Question	N	N (unique)	D	p-value (approx.)
Q1	1602	1039	0.04	0.399
Q2	1602	927	0.05	0.313
Q3	1602	1196	0.04	0.414
Q4	1602	1066	0.03	0.853
Q5	1602	1003	0.04	0.443

Note: This table presents results from two-sample Kolmogorov–Smirnov (KS) test to determine if there are any differences in the distribution of the allocations between the treatment and control groups. The KS test is based on a test statistic that measures the maximum absolute (vertical) difference between the cumulative distribution functions of the two groups. The table shows for each question the maximum absolute difference D. The p-values of this test statistic are based on a five-term approximation of the asymptotic distributions. Given that the KS test is designed for continuous distributions without value ties, the table also presents the number of unique values for each question.
Source: Own calculation based on survey responses.

Table B6: Balancing Tests, Alternative Sample Restrictions

Panel A: Directly Passed Understanding Question

Demographics	Binary Split	N	Point Estimate	Control Mean	RW p-values
Gender	Male vs. Female	1144	0.007	0.5017	0.993
Age	$<$ 45 years vs. $\geq$ 45 years	1154	-0.039	0.6478	0.709
Ethnicity	White/Caucasian vs. Other	1151	0.003	0.1583	0.993
Education	$\leq$ HS vs. $>$ HS	1154	0.022	0.5773	0.945
Employment	(Self)Employed vs. Other	1154	-0.031	0.4141	0.837
Hours	$<$ 35 hours vs. $\geq$ 35 hours	1060	-0.008	0.5028	0.993
Income	$<$ \$40,000 vs. $\geq$ \$40,000	1152	0.013	0.4172	0.987

Panel B: Exclude Response Time Outliers

Demographics	Binary Split	N	Point Estimate	Control Mean	RW p-values
Gender	Male vs. Female	915	0.010	0.4957	0.993
Age	$<$ 45 years vs. $\geq$ 45 years	924	-0.070	0.6858	0.171
Ethnicity	White/Caucasian vs. Other	923	-0.018	0.1592	0.948
Education	$\leq$ HS vs. $>$ HS	924	0.004	0.5987	0.993
Employment	(Self)Employed vs. Other	924	-0.054	0.4289	0.441
Hours	$<$ 35 hours vs. $\geq$ 35 hours	851	-0.004	0.5000	0.993
Income	$<$ \$40,000 vs. $\geq$ \$40,000	922	0.011	0.4286	0.993

Note: This table presents results from the balancing test outlined in equation (1) for different restricted samples. Panel A displays results for the sample of respondents that passed the understanding question of the fairness preference module at the first try. Panel B presents results for a sample that excludes respondents with high (above p90) and low (below p10) response times of the fairness preference module. The table presents point estimates for the coefficient of interest, the mean of the control group (first group in binary split), and associated heteroskedasticity robust p-values adjusted for multiple hypothesis testing using the Romano and Wolf (2016) step-down procedure with 1000 bootstrap draws. Sample sizes differ across demographic characteristics due to the exclusion of responses such as "Prefer not to answer" (see Table 2).
Source: Own calculation based on survey responses.

Table B7: Distribution of Allocations, Kolmogorov–Smirnov Test, Alternative Samples

Panel A: Directly Passed Training Question

Question	N	N (unique)	D	p-value (approx.)
Q1	1154	825	0.05	0.399
Q2	1154	693	0.06	0.306
Q3	1154	904	0.04	0.691
Q4	1154	789	0.03	0.941
Q5	1154	742	0.05	0.385

Panel B: Exclude Response Time Outliers

Question	N	N (unique)	D	p-value (approx.)
Q1	924	700	0.08	0.126
Q2	924	583	0.08	0.103
Q3	924	756	0.07	0.257
Q4	924	655	0.05	0.581
Q5	924	628	0.07	0.253

Note: This table presents results from two-sample Kolmogorov–Smirnov (KS) test for different restricted samples. Panel A displays results for the sample of respondents that passed the training question of the fairness preference module at the first try. Panel B presents results for a sample that excludes respondents with high (above p90) and low (below p10) response times of the fairness preference module. The test determines if there are any differences in the distribution of the allocations between the treatment and control groups. The KS test is based on a test statistic that measures the maximum absolute (vertical) difference between the cumulative distribution functions of the two groups. The table shows for each question the maximum absolute difference D. The p-values of this test statistic are based on a five-term approximation of the asymptotic distributions. Given that the KS test is designed for continuous distributions without value ties, the table also presents the number of unique values for each question.
Source: Own calculation based on survey responses.

Table B8: Allocations, "Real-Person Treatment", Male

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	772	19.77	1139.31	0.909	0.907	0.945
Q2	772	97.57	-2355.06	0.490	0.485	0.857
Q3	772	-478.21	2666.53	0.407	0.396	0.846
Q4	772	-108.54	-4311.73	0.766	0.764	0.945
Q5	772	-691.37	6579.08	0.095	0.104	0.356

Note: This table presents results of the regression analysis outlined in equation (2). We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculations based on survey responses.

Table B9: Allocations, "Real-Person Treatment", Female

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	817	176.46	943.63	0.260	0.249	0.652
Q2	817	-54.76	-2262.16	0.684	0.685	0.791
Q3	817	812.18	1483.75	0.135	0.144	0.471
Q4	817	-384.54	-3893.72	0.296	0.286	0.652
Q5	817	219.52	5812.89	0.560	0.563	0.791

Note: This table presents results of the regression analysis outlined in equation (2). We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculations based on survey responses.

Table B10: Allocations, "Real-Person Treatment", Age below 45

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	693	-120.46	917.07	0.456	0.469	0.893
Q2	693	81.05	-2148.02	0.560	0.556	0.908
Q3	693	132.49	826.70	0.812	0.805	0.915
Q4	693	142.14	-3988.26	0.712	0.692	0.915
Q5	693	-667.37	5687.35	0.103	0.119	0.389

Note: This table presents results of the regression analysis outlined in equation (2). We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculations based on survey responses.

Table B11: Allocations, "Real-Person Treatment", Age 45 and above

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	909	331.63	1121.19	0.041	0.053	0.169
Q2	909	-76.33	-2401.16	0.569	0.559	0.810
Q3	909	467.27	2868.35	0.389	0.383	0.735
Q4	909	-605.29	-4129.73	0.082	0.083	0.273
Q5	909	153.77	6526.30	0.682	0.688	0.810

Note: This table presents results of the regression analysis outlined in equation (2). We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculations based on survey responses.

Table B12: Allocations, "Real-Person Treatment", White / Caucasian Ethnicity

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	1299	136.48	1091.64	0.305	0.278	0.739
Q2	1299	-50.15	-2337.88	0.646	0.648	0.739
Q3	1299	443.18	2294.39	0.320	0.313	0.739
Q4	1299	-509.04	-4020.58	0.075	0.077	0.298
Q5	1299	-228.71	6350.19	0.463	0.465	0.739

Note: This table presents results of the regression analysis outlined in equation (2). We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculations based on survey responses.

Table B13: Allocations, "Real-Person Treatment", Non White / Non Caucasian Ethnicity

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	299	58.40	798.94	0.797	0.782	0.839
Q2	299	260.73	-2121.51	0.221	0.223	0.665
Q3	299	-698.25	915.12	0.392	0.397	0.759
Q4	299	761.15	-4262.26	0.206	0.209	0.665
Q5	299	-323.56	5420.10	0.600	0.623	0.839

Note: This table presents results of the regression analysis outlined in equation (2). We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculations based on survey responses.

Table B14: Allocations, "Real-Person Treatment", High School Education and Below

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	709	148.21	736.75	0.372	0.374	0.820
Q2	709	-29.74	-2207.45	0.838	0.836	0.878
Q3	709	338.06	787.28	0.566	0.564	0.878
Q4	709	-798.36	-3754.04	0.042	0.047	0.184
Q5	709	-268.66	5817.89	0.520	0.508	0.878

Note: This table presents results of the regression analysis outlined in equation (2). We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculations based on survey responses.

Table B15: Allocations, "Real-Person Treatment", More Than High School Education

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	893	53.52	1295.94	0.738	0.723	0.971
Q2	893	47.31	-2375.54	0.718	0.725	0.971
Q3	893	-31.33	3105.79	0.952	0.952	0.971
Q4	893	181.08	-4344.37	0.597	0.605	0.966
Q5	893	-304.67	6498.24	0.414	0.410	0.912

Note: This table presents results of the regression analysis outlined in equation (2). We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculations based on survey responses.

Table B16: Allocations, "Real-Person Treatment", (Self-) Employed

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	992	104.00	1063.81	0.468	0.487	0.903
Q2	992	-26.56	-2184.63	0.825	0.826	0.903
Q3	992	577.65	1482.95	0.238	0.228	0.678
Q4	992	-228.93	-3958.98	0.489	0.500	0.903
Q5	992	-217.43	5883.71	0.538	0.564	0.903

Note: This table presents results of the regression analysis outlined in equation (2). We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculations based on survey responses.

Table B17: Allocations, "Real-Person Treatment", Unemployed/Student/Retiree

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	610	129.04	997.67	0.511	0.510	0.939
Q2	610	53.03	-2475.35	0.746	0.740	0.939
Q3	610	-336.65	2901.99	0.609	0.600	0.939
Q4	610	-316.62	-4248.97	0.444	0.484	0.939
Q5	610	-304.58	6654.78	0.500	0.507	0.939

Note: This table presents results of the regression analysis outlined in equation (2). We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculations based on survey responses.

Table B18: Allocations, "Real-Person Treatment", Working Less Than 35 Hours/Week

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	712	-51.47	1160.55	0.774	0.778	0.928
Q2	712	66.53	-2362.80	0.649	0.634	0.928
Q3	712	-388.03	2460.73	0.521	0.524	0.928
Q4	712	-243.60	-4156.77	0.522	0.528	0.928
Q5	712	-466.87	6390.95	0.266	0.249	0.747

Note: This table presents results of the regression analysis outlined in equation (2). We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculations based on survey responses.

Table B19: Allocations, "Real-Person Treatment", Working 35 Hours/Week or More

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	756	214.37	1078.48	0.198	0.179	0.580
Q2	756	12.33	-2261.43	0.930	0.930	0.930
Q3	756	564.42	1923.83	0.320	0.286	0.740
Q4	756	-178.92	-3953.36	0.637	0.633	0.920
Q5	756	-216.74	6108.81	0.589	0.573	0.920

Note: This table presents results of the regression analysis outlined in equation (2). We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculations based on survey responses.

Table B20: Allocations, "Real-Person Treatment", Earnings Less Than 40000

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	941	46.18	997.80	0.753	0.727	0.977
Q2	941	-5.75	-2314.82	0.963	0.965	0.995
Q3	941	43.72	1641.26	0.932	0.939	0.995
Q4	941	-260.73	-4050.91	0.450	0.426	0.927
Q5	941	-171.49	6122.00	0.630	0.630	0.970

Note: This table presents results of the regression analysis outlined in equation (2). We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculations based on survey responses.

Table B21: Allocations, "Real-Person Treatment", Earnings 40000 or More

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	658	213.33	1095.28	0.258	0.257	0.732
Q2	658	36.73	-2281.88	0.813	0.828	0.859
Q3	658	391.29	2666.52	0.521	0.520	0.859
Q4	658	-266.91	-4085.23	0.495	0.496	0.859
Q5	658	-403.93	6281.06	0.368	0.381	0.791

Note: This table presents results of the regression analysis outlined in equation (2). We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculations based on survey responses.

Table B22: Allocations, "Real-Person Treatment", 5 Questions

Question	N	Point Estimate	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	534	248.68	0.211	0.203	0.562
Q2	534	156.33	0.377	0.389	0.717
Q3	534	-543.47	0.434	0.432	0.717
Q4	534	-102.52	0.823	0.803	0.803
Q5	534	-964.28	0.053	0.058	0.223

Note: This table presents results of the regression analysis outlined in equation (2). We present point estimates for the coefficient of interest β^j, and associated heteroskedasticity robust model p-values, as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications. We focus on those respondents that answer exactly five questions.
Source: Own calculations based on survey responses.

Table B23: Allocations, "Real-Person Treatment", 7 Questions

Question	N	Point Estimate	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	536	-42.11	0.830	0.835	0.991
Q2	536	115.69	0.479	0.462	0.942
Q3	536	-464.22	0.480	0.468	0.942
Q4	536	-454.02	0.299	0.315	0.864
Q5	536	-114.90	0.803	0.811	0.991
Q6	536	-17.16	0.965	0.957	0.991
Q7	536	-109.54	0.434	0.433	0.942

Note: This table presents results of the regression analysis outlined in equation (2). We present point estimates for the coefficient of interest β^j, and associated heteroskedasticity robust model p-values, as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications. We focus on those respondents that answer exactly seven questions.
Source: Own calculations based on survey responses.

Table B24: Allocations, "Real-Person Treatment", 9 Questions

Question	N	Point Estimate	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	532	137.73	0.510	0.514	0.961
Q2	532	-248.76	0.131	0.133	0.571
Q3	532	1653.34	0.017	0.013	0.121
Q4	532	-209.97	0.639	0.636	0.961
Q5	532	291.45	0.550	0.560	0.961
Q6	532	43.65	0.913	0.911	0.961
Q7	532	122.00	0.362	0.381	0.913
Q8	532	397.68	0.537	0.518	0.961
Q9	532	832.08	0.189	0.192	0.684

Note: This table presents results of the regression analysis outlined in equation (2). We present the number of respondents, the number of rejected hypotheses according to heteroskedasticity robust model p-values, resample p-values, and p-values adjusted for multiple hypothesis testing. We focus on those respondents that answer exactly nine questions.
Source: Own calculations based on survey responses.

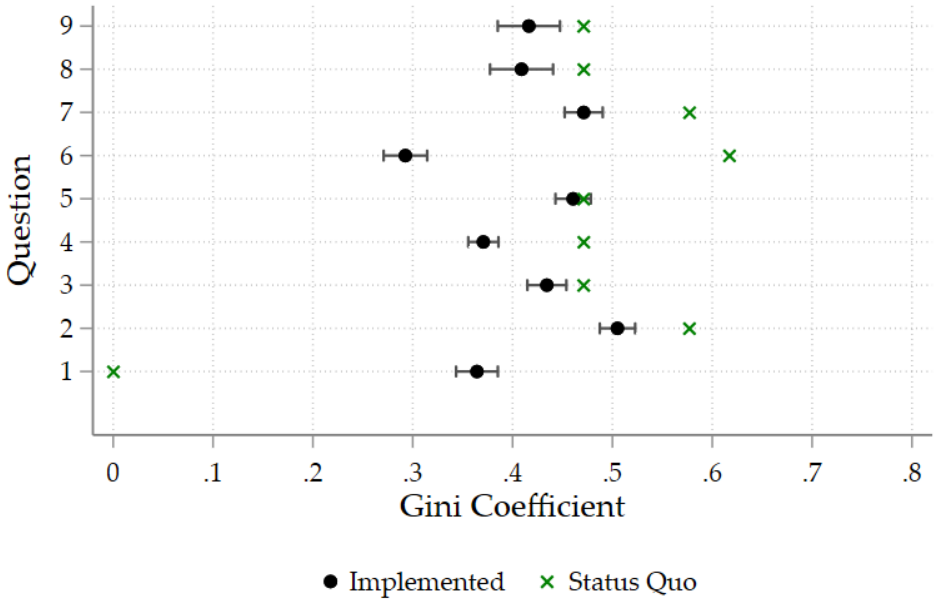
Table B25: Preferences, Response Time (Min.), "Real-Person Treatment", No Limit

Question	N	Point Estimate	Control Mean	Model p-value	Resample p-value	Romano-Wolf p-value
Q1	1602	-0.28	1.00	0.246	0.276	0.758
Q2	1602	0.03	0.46	0.490	0.513	0.758
Q3	1602	-0.03	0.50	0.476	0.483	0.758
Q4	1602	-0.02	0.37	0.344	0.386	0.758
Q5	1602	-0.04	0.36	0.242	0.280	0.758

Note: This table presents results of the regression analysis outlined in equation (2) using the response time in minutes as the dependent variable. We present point estimates for the coefficient of interest β^j, the mean of the control group, and heteroskedasticity robust uncorrected analytical p-values (model p-values), uncorrected bootstrapped p-values (resample p-values), as well as p-values adjusted for multiple hypothesis testing using 1000 bootstrap replications.
Source: Own calculation based on survey responses.

C Nature of Fairness Preferences

Figure C4: Gini Coefficient, Treatment Group



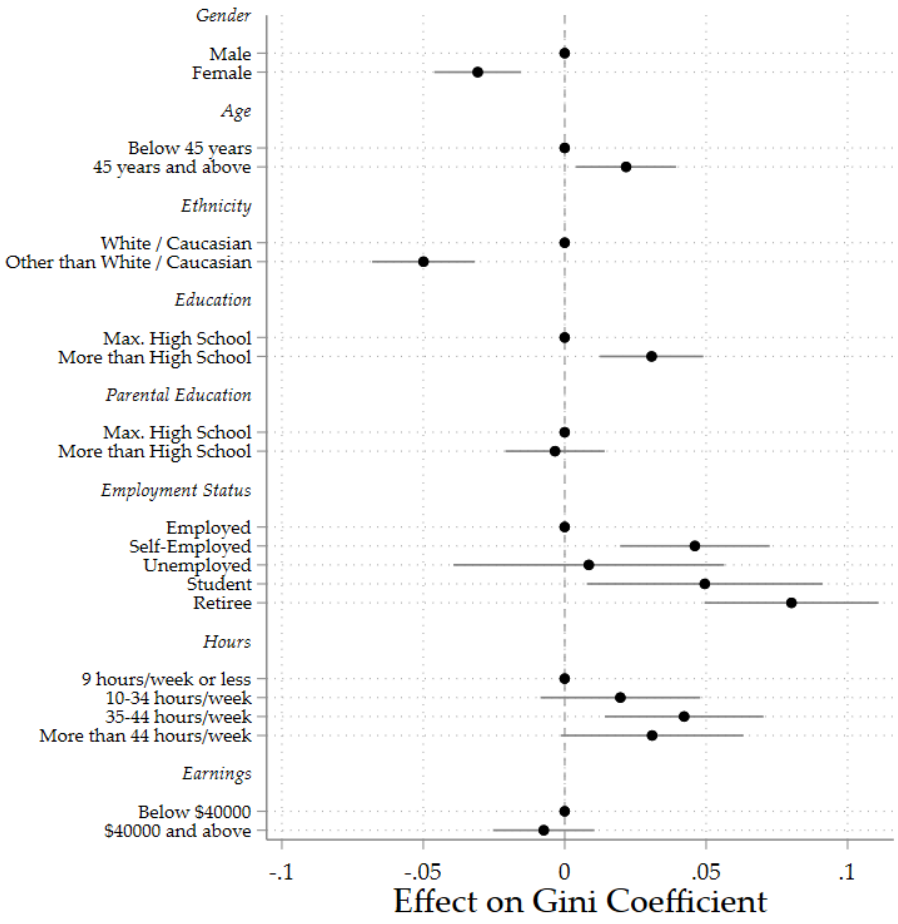
Note: This figure compares implemented Gini coefficients (gray dots) to initial Gini coefficients (green crosses) in each vignette. Gray bars indicate 95 percent confidence intervals. Gini coefficients are calculated at the respondent level as $\frac{|x-y|}{x+y}$ where x (y) is the amount allocated to vignette person A (B). We focus exclusively on respondents in the treatment group, i.e., those respondents who were informed about the real-world consequences of their choices.
Source: Own calculations based on survey responses.

Table C26: Preference Types, Treatment Group

	Share Egalitarians (%)			Share Libertarians (%)		
Max. abs. difference (pp)	2	5	10	2	5	10
Allow for 0 inconsistent answers	0.72	1.07	2.32	5.74	6.27	8.13
Allow for 1 inconsistent answers	1.07	**2.17**	5.58	7.66	**9.04**	12.95
Allow for 2 inconsistent answers	1.49	3.82	10.49	10.36	13.28	22.37

Note: This table presents shares of egalitarians (libertarians) according to consistent choices in all questions of the baseline wave. We also vary the leniency of the classification by allowing for 0, 1, 2 answers that are inconsistent with egalitarian (libertarian) choices. In the baseline (highlighted estimates), we allow for a deviation of +/-5 percentage points and inconsistent choices in one question only. We focus exclusively on respondents in the treatment group, i.e., those respondents who were informed about the real world consequences of their decisions.
Source: Own calculations based on survey responses.

Figure C5: Gini Coefficient and Respondent Demographics



Note: This figure displays the coefficients from a regression that regresses the implemented Gini coefficient on various demographic characteristics, i.e., gender, age, ethnicity, education, parental education, employment status, working hours, and labor market earnings of the respondent. The graph includes the base level, the point estimate for the other level(s), as well as 95 percent confidence intervals. We focus exclusively on respondents in the control group, i.e., those respondents who faced hypothetical scenarios.
Source: Own calculations based on survey responses.

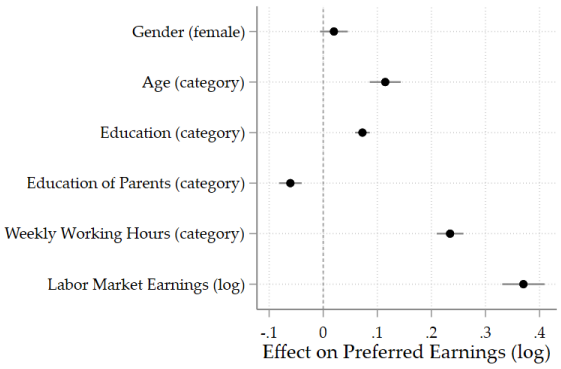
Table C27: Fairness Preferences, Regression Table

Vignette Person Char.	Log (Preferred Earnings)		Preferred Share (%)		Log(Preferred Ratio)	
Female (binary)	0.009	0.010	-0.149	-0.136	-0.009	-0.008
	(0.511)	(0.456)	(0.738)	(0.761)	(0.736)	(0.756)
Age (binary)	0.154	0.152	3.409	3.390	0.202	0.201
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education (cat. 1,2,3)	0.091	0.089	2.411	2.394	0.138	0.138
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Parental Education (cat. 1,2,3)	-0.047	-0.045	-3.872	-3.851	-0.212	-0.211
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weekly Working Hours (cat. 1,2,3,4)	0.233	0.234	7.950	7.956	0.436	0.437
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log(Earnings)	0.380	0.379	.	.	.	.
	(0.000)	(0.000)				
Earnings Share	.	.	4.136	4.132	.	.
			(0.000)	(0.000)		
Log(Ratio)	.	.	.	.	0.428	0.427
					(0.000)	(0.000)
Constant	4.149	4.163	17.008	17.025	-0.694	-0.695
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sum of Earnings FE	Yes	Yes	Yes	Yes	Yes	Yes
Respondent FE	No	Yes	No	Yes	No	Yes
Number of Observations	10980	10980	10980	10980	10980	10980

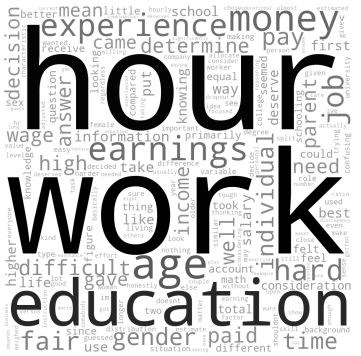
Note: This table presents regression results from equation 4 where we regress the outcome variable of interest on linearized categorical information on different vignette person characteristics (column 1). The regression further controls for the total amount of earnings in a person pair and, as a robustness test, for respondent fixed effects. The table provides regression results for three outcome variables of interest, i.e., the log value of preferred earnings (column 2-3), the preferred earnings share (column 4-5), and the log value of the preferred ratio (column 5-6). We focus exclusively on respondents in the control group, i.e., those respondents who faced hypothetical scenarios. Point estimates are shown with respective p-values based on heteroskedasticity robust standard errors clustered at the respondent level in brackets. To keep the number of observations constant across columns, we focus on non-zero earnings allocations across all specifications.
Source: Own calculations based on survey responses.

Figure C6: Fairness Preferences, Importance of Profile Characteristics, Treatment Group

(a) Regression Analysis



(b) Text Analysis



Note: This figure displays how respondents take the displayed characteristics in a profile into account for their distributional choices. Figure C6a displays the coefficients from a regression analysis following equation (4) that regresses the preferred earnings (log) on linearized categorical information on the person's characteristics. The regression further controls for the total amount of earnings in a person pair. The figure shows point estimates and heteroskedasticity robust 95 percent confidence intervals. Figure C6b displays a word cloud from a text analysis of the open-ended question in the fairness preference module. At the end of the module, people were asked about how they came up with their distributional choices and to describe their reasoning in their own words. Based on the text corpus from the open-ended answers, we used natural language processing techniques to rank the frequency of specific terms. We focus exclusively on respondents in the treatment group, i.e., those respondents who were informed about the real-world consequences of their choices.
Source: Own calculation based on survey responses.