

Kosfeld, Michael; Sharafi, Zahra; González, Maíra Sontag; Zou, Na

**Working Paper**

## Measuring Economic Preferences with Surveys and Behavioral Experiments

CESifo Working Paper, No. 11631

**Provided in Cooperation with:**

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

*Suggested Citation:* Kosfeld, Michael; Sharafi, Zahra; González, Maíra Sontag; Zou, Na (2025) : Measuring Economic Preferences with Surveys and Behavioral Experiments, CESifo Working Paper, No. 11631, CESifo GmbH, Munich

This Version is available at:

<https://hdl.handle.net/10419/314670>

**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.*



# Measuring Economic Preferences with Surveys and Behavioral Experiments

*Michael Kosfeld, Zahra Sharafi, Maíra Sontag González, Na Zou*



## **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](mailto: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](http://www.SSRN.com)
- from the RePEc website: [www.RePEc.org](http://www.RePEc.org)
- from the CESifo website: <https://www.cesifo.org/en/wp>



# Measuring Economic Preferences with Surveys and Behavioral Experiments

## Abstract

Developing reliable and practicable measures of economic preferences is a crucial task for empirical economic research with high value for both theory and applications. Here, we present results from a first comprehensive “behavioral validation analysis” of the Global Preference Survey Module (GPS) and the corresponding Preference Survey Module (PSM) developed by Falk et al. (2018, 2023) that have been widely used for the measurement and analysis of economic preferences on a global scale. Our key questions are how well GPS and PSM modules explain behavior in incentivized choice experiments in other countries than in the original validation in Germany, and to what extent survey items and modules developed from behavioral experiments in different countries and cultures resemble one another. Our current results, which are based on experiments in three very diverse countries—China, Iran, and Kenya—show that many GPS and PSM survey items predict behavior in incentivized choice experiments, but coefficients vary and are not always sizable. Quantitative items, which are based on hypothetical choice experiments, are consistently selected into survey modules, whereas the best qualitative items differ between countries. At the same time, the contribution in terms of explanatory power of these latter items is comparably lower. Our analysis provides a first empirical basis for the development of survey modules that reliably predict behavior in incentivized choice experiments and real-life situations across diverse countries and contexts. Additional results, including principal component analysis and prediction of real-life behavior, highlight important gaps that warrant further investigation in future research.

JEL-Codes: C830, C910, D010.

Keywords: economic preference, measurement, survey, experiment.

*Michael Kosfeld*  
*Faculty of Economics and Business*  
*Goethe University Frankfurt / Germany*  
*kosfeld@econ.uni-frankfurt.de*

*Zahra Sharafi\**  
*Frankfurt School of Finance and Management*  
*Frankfurt am Main / Germany*  
*z.sharafi@fs.de*

*Maíra Sontag González*  
*Faculty of Economics and Business*  
*Goethe University Frankfurt / Germany*  
*sontaggonzalez@econ.uni-frankfurt.de*

*Na Zou*  
*University of Bath School of Management*  
*Bath / United Kingdom*  
*nz464@bath.ac.uk*

\*corresponding author



January 9, 2025

We thank Sen Geng, Johannes Haushofer, Rosa Hendijani, and the Busara Center for Behavioral Economics, Inc. for their help and support in conducting the lab experiments. We are grateful to Peter Andre, Anke Becker, Volker Benndorf, Jonathan Chapman, Felix Chopra, Julian Detemple, Thomas Dohmen, Armin Falk, Andreas Grunewald, Florian Hett, David Poensgen, Frederik Schwerter, Erik Snowberg, Ferdinand von Siemens, and Matthias Wibral, as well as seminar participants at Goethe University Frankfurt, Frankfurt School, DICE, ESA Helsinki 2024, and the Behavioral Measurement Conference 2024 for helpful comments and suggestions. We gratefully acknowledge financial support from the Diligentia Stiftung. IRB approval was received from the Joint Ethics Committee of Goethe University Frankfurt and Johannes Gutenberg University Mainz.



# 1 Introduction

According to economic theory, individual preferences, together with beliefs, determine (rational) human behavior. Prime examples include preferences about risk, time, and welfare outcomes of second or third parties, i.e., so-called “social” preferences. Variation in these preferences has been documented to predict human decision-making and be associated with important life outcomes across a wide range of social and economic domains. For example, risk preferences are associated with sorting into occupations with different earnings risk (Bonin et al., 2007), performance pay (Caliendo et al., 2009; Grund and Sliwka, 2010), geographic mobility (Jaeger et al., 2010), self-employment (Dohmen et al., 2011), or investment in risky assets (Dohmen et al., 2011). Time preferences are found to be associated with health outcomes (Kirby and Petry, 2004; Chabris et al., 2008), education outcomes (Golsteyn et al., 2014; Backes-Gellner et al., 2021), savings (Sutter et al., 2013), and credit card borrowing (Meier and Sprenger, 2010; Golsteyn et al., 2014). Further, social preferences predict success in collective action (Rustagi et al., 2010; Carpenter and Seki, 2011; Kosfeld and Rustagi, 2015), the development and maintenance of social capital (Cárdenas et al., 2013), and self-selection into public service (Friebel et al., 2019). Together these results elucidate the importance of a reliable, i.e., accurate and precise, measurement of economic preferences.

Traditionally, economists follow the “revealed-preference approach” (Samuelson, 1938; Sen, 1973) by which preferences can be measured—are revealed—through individual choices made in controlled decision-making situations. By varying elements and characteristics of the choice set, different types of preferences can be measured (e.g., ambiguity vs. risk, reciprocity vs. inequity aversion, etc.). While this approach has established incentivized choice experiments as a key methodological benchmark for measuring economic preferences, its relative complexity and resource intensity impose significant limitations. This is especially the case when it comes to the measurement of preferences in large, representative populations.

To address these limitations, Falk et al. (2018, 2023) develop the Global Preference Survey (GPS) as an innovative survey instrument to measure key economic preferences (risk, time, social) on a global scale. The data collected and summarized in Falk et al. (2018) include observations from around 80,000 individuals in representative population samples across 76 countries. Survey items in the GPS are derived from prior incentivized choice experiments conducted with about 400 students at the University of Bonn (Falk et al., 2023). For each type of preference, students participated in corresponding choice experiments and answered a large battery of survey questions. Subsequently, the two questions—one quantitative (based



on a hypothetical choice scenario) and one qualitative (based on self-assessment)—that best explain the observed variation in behavior in the choice experiments are selected for the so-called Preference Survey Module (PSM) (Falk et al., 2023). For the GPS module, some of these questions are further refined or replaced to adapt the instrument for a global field implementation. Results in Falk et al. (2018) demonstrate the validity of their preference measures by showing significant correlations, both across and within countries, with relevant economic outcomes and plausible influencing factors. Subsequent work has produced additional compelling evidence (e.g., Lades et al., 2021; Sunde et al., 2022; Hanushek et al., 2022).

Still, an important open question is whether survey items in the GPS or the PSM explain behavior in incentivized choice experiments in other countries than in the original validation in Germany. In this paper, we fill this gap. We conduct the first full validation analysis of Falk et al. (2023) across multiple countries using the exact same procedures as in the original study. Our analysis includes student samples from diverse regions of the world—currently China, Iran, and Kenya. We are in the process of collecting additional data from Colombia and the United States in 2025. The approach allows us to assess the cross-cultural reliability and predictive power of the employed preference measures across varied cultural contexts. In particular, we examine whether survey items in the PSM and GPS modules are significantly associated with behavior in incentivized choice experiments in these countries. Moreover, using the same item selection procedure as in Falk et al. (2023), we test whether the items selected, both for each country separately and combined, are identical to those in the PSM from Germany.

Our current results are as follows. First, we find that many GPS and PSM survey items—both quantitative and qualitative—perform well in predicting behavior in incentivized choice experiments, but coefficients vary and are not always sizable. For negative reciprocity in most countries and several preferences in Kenya, performance is found to be relatively poorer compared to the other preferences. Second, using the same item selection procedure as in Falk et al. (2023), we identify the two survey items with the highest explanatory power for each preference. While the qualitative items vary between countries, the quantitative items are consistently selected across and within countries, highlighting them as reliable predictors for the behavior in experiments. Given their significantly higher power in explaining behavior, the loss due to the observed variation in qualitative items between countries is found to be rather limited.

These results speak to the few existing studies that have tried to replicate correlations between GPS measures and behavior in experiments for selected countries and preferences. For instance, Bauer et al. (2020) find in a sample of low-income households in Kenya that only the quantitative GPS items predict



corresponding behavior in incentivized choice experiments while qualitative items perform poorly. Kosfeld and Sharafi (2023) observe similar patterns with university students in Iran. However, with only about 100 participants each, it is difficult to rule out that differences in the performance of quantitative and qualitative items in these two studies are driven by the fact that qualitative items may simply be noisier. Next, neither study considers the full set of GPS and PSM items for all economic preferences. Kosfeld and Sharafi (2023), for example, only look at social preferences. More recently, Chapman et al. (2024) examine the validity of qualitative self-assessment questions in a general population with large sample sizes in the United States, casting doubt on using them as reliable indicators of target preferences. Our findings contribute to these different pieces of evidence by unveiling the full picture of how GPS and PSM measures replicate in behavioral experiments with student samples across countries. In a nutshell, we find that there is important variation but by and large the measures do better than expected—considering the poor performance in other replication studies. This is particularly due to the performance of quantitative survey items—a finding that is well in line with these other studies.

We address a number of additional questions in our analysis. First, we explore different ways to refine the survey module to enhance its adaptability and portability in varied cultural contexts for future research. Using behavioral measures still as the benchmark for eliciting preferences, we develop alternative survey modules, either single- or multiple-item. In developing these modules, we select items that demonstrate an association with observed experimental behavior in as many countries as possible. This approach ensures that the modules remain effective when applied by researchers in diverse contexts, thus enhancing their reliability and adaptability. We also provide an evaluation matrix comparing their parsimony and accuracy. Our goal here is to provide researchers with relevant information to make a research-specific decision on which module to use in their own research. Next, taking one step back without considering experimental behavior as the object to be approximated, we conduct a principal component analysis on responses to all survey questions to derive the latent factors underlying a given economic preference. For each preference, we identify four to five components, in which (hypothetical as well as incentivized) experimental choice scenarios are always clustered into one component.

Finally, given the well-established association between economic preferences and real-life behavior documented in the literature, we investigate which methodology—whether incentivized choice experiments or survey questions—provides a better explanation of outside-lab behaviors. We find that while preferences elicited via survey questions are consistently and significantly correlated with real-life behavior in the do-



mains of health, finance, and social interactions, choice-based measures are only weakly correlated with real-life behavior. An important caveat is that these outside behaviors are based on self-reports, i.e., are elicited in the same way as the preference survey questions.

Altogether, this paper makes three important contributions to the literature. First, we document that most GPS and PSM survey items are found to be significantly correlated with behavior in incentivized choice experiments across countries beyond Germany. Prior to our study, the predictive power of these modules in diverse cultural contexts was unknown. We build on existing studies, which often relied on partial replication or small sample sizes, by conducting the first full replication with larger student samples drawn from multiple countries representing various continents and cultures. Second, this extensive dataset allows us to develop alternative measures and modules that are informed by the observed variation in behavior in incentivized choice experiments across cultural contexts. Finally, our study provides a first comparison of two primary methods for measuring economic preferences—incentivized choices and survey questions—regarding their ability to predict real-life behaviors. Ultimately, the key question is, of course, how do we best measure economic preferences. Our suggestive evidence highlights a number of critical gaps that warrant further investigation in future research. Additionally, our research lays the groundwork for other researchers to join our initiative in the future, e.g., by utilizing the study materials we develop in their own laboratories across various countries, thus advancing the research agenda toward a comprehensive validation and development of truly global measures of economic preferences.

## **2 Study Design and Procedures**

We implemented the complete study package described in Falk et al. (2023) to measure economic preferences regarding risk taking, time discounting, trust, altruism, positive reciprocity, and negative reciprocity, following the exact same procedures as in the original study.

To find local labs, we screened active economic laboratories in China, Iran, and Kenya and selected the most compatible one in each country, ensuring it met the standards of typical econ labs and could comply with the particular aspects of our study.<sup>1</sup> In each country, we recruited students from a public, comprehensive university comparable to the University of Bonn in Germany. In Iran we conducted our study at the Business Research Lab at the University of Tehran in April and May 2023; In China at the

---

<sup>1</sup>In line with the original study, we required participants to have no prior experience in any economic experiments, to come to the lab twice at the exact time one week apart, and depending on their choices in the experiment, to be paid in one year's time.



FEEL lab at Xiamen University in October 2023; and in Kenya at the Busara Lab with students from the University of Nairobi in November and December 2023. In total, 1167 students (395 from China, 389 from Iran and 383 from Kenya) participated in our study.

For each preference, participants are asked a set of both quantitative and qualitative survey questions. Following the standard practice of translation and back-translation by native or bilingual speakers (Brislin, 1970), we created the English master version based on the original German codes in zTree, and translated them into Mandarin and Farsi. As the official language in Kenya is English, we conducted our study in Kenya in English. All experiments and surveys were programmed in oTree (Chen et al., 2016).

To minimize spillovers between experimental and survey measures, we followed the procedures of Falk et al. (2023) and invited each participant twice, with one session containing experiments on risk and time preferences matched with survey items on social preferences, and another session containing experiments on social preferences matched with survey items on risk and time preferences. Both sessions were one week apart. To mitigate potential order effects, the order of sessions was randomized such that half of the participants started their first session with experiments on risk and time preferences, whereas the other half started with experiments on social preferences. Table 1 provides an overview of the study design.

Group	Session	Week	Payment
1	A: Experiments on risk and time preferences; survey on social preferences	T	T and T+12 months
	B: Experiments on social preferences; survey on risk and time preferences	T+1	T+1
2	B: Experiments on social preferences; survey on risk and time preferences	T	T
	A: Experiments on risk and time preferences; survey on social preferences	T+1	T+1 and T+1+12 months

Table 1: **Schedule.** Time schedule for each week, group, and experiment.

We calibrated monetary incentives for each country using prices of a comparable food item in the spirit of the Big Mac Index to adjust for differences in purchasing power between countries (Parsley and Wei, 2007). In the original study conducted in Germany in winter 2010/2011, every 100 experimental points were worth 0.8 euros. Adjusting for inflation and differences in local purchasing power, every 100 experimental points in our study were equivalent to 4 Yuan, 150,000 Iranian Rial, and 30 Kenyan Shilling in China, Iran, and



Kenya, respectively. The resulting average earnings over both sessions, including show-up fees, amounted to 52, 39 and 20 euros in China, Iran, and Kenya, respectively. The amounts correspond well in real terms to the average earnings in Falk et al. (2023). On average, a session lasted about 100 minutes. Table A1 in Appendix A.1 presents summary statistics of participants.

## 2.1 Incentivized Choice Experiments

We made no changes to the incentivized choice experiments or the measurement method implemented by Falk et al. (2023). We retained the games, instructions, order, and the perfect-stranger-matching in games involving social or strategic interaction as described in the original study. To reduce measurement errors, two variations of an experiment were conducted to measure risk taking, time discounting, trust, and positive reciprocity. These versions share the same structure but have slightly different payoffs. The experimental measure for a given preference is then calculated by averaging the two choices in each version (see Table D2 in Appendix D).

Concerning time discounting experiments in which participants have to choose between a payment today and a payment in 12 months, we deviated from the payment method used in Falk et al. (2023) of sending future payments in cash by regular mail. Instead, we utilized established payment methods in local labs, in particular digital transfers via known platforms in China, Iran, and Kenya. Participants were assured they would receive their delayed payments with the institutional guarantee offered by local labs. They were further informed that future payments will be inflation-adjusted. Finally, and importantly, we added control questions to the experiments to ensure participants understand the games, and to control for their level of comprehension in our analysis.

**Risk Preferences** - For risk taking, participants completed two multiple price lists, each containing 21 choices between a lottery that yields 1000 points with 50% chance and 0 points otherwise, and a safe payment option that increased with each subsequent decision. Risk taking is measured by averaging over the switching rows (i.e., the rows in which participants switched from preferring the lottery to the safe payment) in these two multiple price lists.

**Time Preferences** - Similarly, for time discounting, participants completed two multiple price lists, each containing 25 choices between a fixed amount of immediate payment (today) and a payment in 12 months which increased with each subsequent decision. Time discounting is measured by averaging over the switching rows (i.e., the rows in which participants switched from preferring a payment today to a



payment in 12 months) in these two multiple price lists.

**Trust** - Trust is measured with the first-mover behavior in two different versions of the investment game (Berg et al., 1995), where the amount sent by the first mover is doubled or tripled. Each participant takes on the role of the first and second mover in each version, resulting in four investment games for each participant. The average amount sent as a first mover in both versions determines the individuals' willingness to trust strangers.

**Altruism** - Altruism is measured by a dictator game in which a charitable organization serves as the recipient. Each country's participants were given a list of well-established international and domestic charitable organizations in that country. Participants could designate a charity organization to which they wanted the money to be donated. We elicited altruism with the amount participants transferred to charity.<sup>2</sup>

**Positive Reciprocity** - Positive reciprocity is measured with the second-mover behavior in the two above-mentioned investment games. We used the strategy method to elicit the second-mover behavior and took the average of these decisions in each version of the game. The average from the two games is then used to measure positive reciprocal inclination.

**Negative Reciprocity** - Finally, two distinct experimental games are employed to measure negative reciprocity: The ultimatum game (Güth et al., 1982) and the prisoner's dilemma game with punishment (Falk et al., 2005). Participants took part in two ultimatum games, one in which they were the sender and the other in which they were the receiver, with the sequence selected randomly. We used participants' minimum acceptable offer in the ultimatum game as the first indicator for negative reciprocity. Subsequently, individuals participated in a prisoner's dilemma game with punishment. We used participants' investment in punishing unilateral defection of their opponents as the second indicator for negative reciprocity. To obtain the measure of negative reciprocity, we first standardized these two indicators and then took the average.

Summary statistics of participants' choices and answers across countries are presented in Tables A3 to A8 in Appendix A.3. Experimental instructions including control questions are reported in Appendix E.1. We document the share of participants who answered all control questions correctly in Appendix A.2. As can be seen, there is substantial heterogeneity across countries and preferences. Specifically, participants from China showed the highest rate of correct answers across most experiments, with particularly high rates in the risk preference (97%) and prisoner's dilemma (98%) experiments. Participants in Iran showed the

---

<sup>2</sup>Participants indicated the amount they wanted to transfer to the charities, and we, the experimenters, made the transactions on their behalf. All receipts were uploaded to protected cloud folders so that participants in each country could verify that their donations had been made successfully.



highest rate of correct answers in the time preference (95%) and donation game (90%) experiments but revealed considerably lower rates in the ultimatum game (61%) and the prisoner’s dilemma (72%). Overall, participants in Kenya showed the lowest average rate, especially in the risk preference (67%) and ultimatum game (56%) experiments, but showed high rates in the time preference experiments (94%). While these rates document a significant share of participants who did not answer control questions correctly, our procedures tried to ensure that in the end, all participants understood the instructions of each experiment before taking decisions. In particular, every participant was provided the correct answer to each control question in case they did not answer it correctly beforehand. Next, if a participant showed particular trouble, we came to the participant and explained instructions in more detail in person. Importantly, as we show below, our results do not depend on whether we include all participants or only those who answered control questions correctly.

## **2.2 Survey Items on Preferences and Real-life Behavior**

While we kept the incentivized choice experiments and their measurement methods unaltered, we made some important adjustments to the survey items of the original study. It is worth noting that the original study was already very comprehensive, containing 188 qualitative and quantitative survey items. These items came from a variety of sources: Some were taken from existing surveys such as the German Socio-Economic Panel Study (SOEP), the National Longitudinal Study of Youth (NLSY), and the World Values Survey (WVS), others were drawn from past research (Weber et al., 2002), or were created and added by the authors (Falk et al., 2023). For each preference, respondents answer triangular questions about their self-assessment, how they compare themselves to others, and how they perceive others to view them. Along with multiple qualitative questions in several related domains of a given preference, participants also respond to quantitative questions that are typically hypothetical versions of the experiments.

The original set of items contains questions with similar content, questions that have been shown to not capture individual variance, that require respondents to possess particular characteristics (e.g., employment), refer to a context that is in conflict with legal or social norms in some countries (e.g., gambling) or a situation that appears ambiguous (e.g., who is treated unfairly by others, who is the transgressor of a norm, who gets punished, etc.). Some questions also measure two distinct preferences simultaneously (e.g., positive as well as negative reciprocity), which makes it difficult to interpret the results. We, therefore, refined the survey battery to obtain a final set of 120 qualitative and quantitative survey items. In Appendix E.2 we provide



a detailed description of our iteration process and display all questions that remained as well as those that were excluded (Appendix E.3). All GPS and PSM items are of course still included in our study (see Table D1 in Appendix D). Summary statistics for all survey questions across countries are presented in Tables A3 to A8 in Appendix A.3.

To compare the performance of behavior- versus survey-based preference measures in explaining outside-lab behavior, we added some new questions that elicit information about participants’ real-life behavior in domains like health, social engagement, and financial decision-making (Anderson and Mellor, 2008; Dohmen et al., 2009, 2011; Lades et al., 2021; Sutter et al., 2013). To avoid any interference with participants’ responses to the main body of experiments and survey questions, we programmed these new questions as a separate module at the end of the last session. These additional questions provide a unique opportunity to provide a first assessment of the relative performance of the two different methodologies in predicting economic behavior outside the lab. Finally, we also elicited demographics such as age, gender, cognitive ability, and socioeconomic status at the end of the study.

### **3 Comparing the Performance of Survey Modules in Different Countries**

#### **3.1 Global Preference Survey (GPS)**

In this section, we investigate whether survey items in the GPS explain behavior in incentivized choice experiments in countries beyond the original validation in Germany. We begin by assessing whether GPS questions can significantly predict behavior in the experiments in each country, as well as the combined data from all replication countries.

For each preference, we run OLS regressions of the standardized experimental measure on the standardized corresponding GPS survey items. Figure 1 illustrates the results by preference and country. The first coefficient on the top in each panel presents results from the original study in Germany, serving as a benchmark.<sup>3</sup> It is calculated based on data from the replication package published by Falk et al. (2023). The subsequent coefficients correspond to China, Iran, and Kenya, respectively. The final two coefficients in each panel show results for all replication countries combined, both with and without country fixed effects. The complete regressions are reported in Table B1 in Appendix B.

---

<sup>3</sup>GPS questions for negative reciprocity were later modified and therefore did not appear in the original validation study.



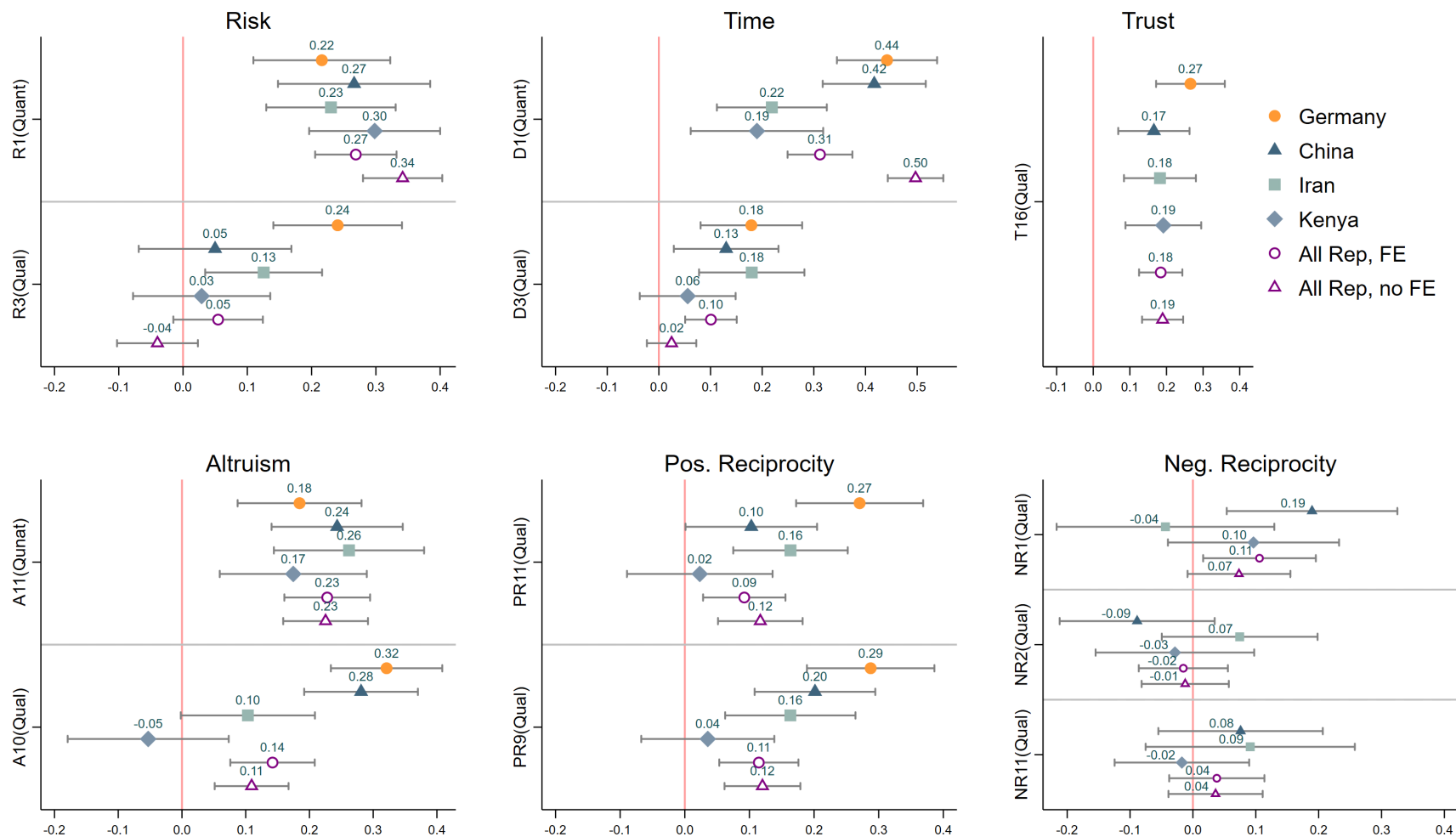


Figure 1: **Plot of regression coefficients of GPS items.** Coefficients from OLS regressions with 95% confidence intervals. For negative reciprocity, the original German sample does not have GPS items NR1 and NR2. R1: switching row from staircase measure; R3: willingness to take risks. D1: switching row from staircase measure; D3: willingness to wait. T16: I assume that people have only the best intentions. A11: Hypothetical donation game; A10: Share when it comes to good causes. PR11: When someone does me a favor, I am willing to return it; PR9: Thank-you gift in the lost-in-the-city scenario. NR1: Punish someone who treats you unfairly; NR2: Punish someone who treats others unfairly; NR11: Take revenge on the first occasion.



At first glance, the results in Figure 1 suggest that GPS items seem to predict behavior in the incentivized choice experiments rather well, with the exception of negative reciprocity and a comparably better overall performance of the quantitative survey items. Still, the magnitude of coefficients in the different countries and in all countries combined varies. Results for the quantitative items—R1, D1, and A11—show relatively strong and statistically significant coefficients across all countries, making them reliable predictors of experimental choices in the lab. The coefficients of these items, particularly those related to risk taking and altruism, are remarkably similar across countries. This aligns with prior evidence regarding the performance of quantitative survey items (Bauer et al., 2020; Kosfeld and Sharafi, 2023). Instead, the qualitative items generally show more variation in their predictive power, with the trust question (T16) being an exception. However, the overall performance of these survey items appears considerably better than in the aforementioned prior studies.<sup>4</sup>

Figure 1 identifies whether the 95% confidence intervals of the effect sizes in each country include the original effect size observed in the German sample (reported at the top of each panel). This replicability assessment, outlined by Camerer et al. (2016), provides further insights into the performance of different questions and allows for a comparison of their replicability. Notably, the quantitative questions of risk and altruism meet the 95% confidence interval criterion in all countries separately and combined.<sup>5</sup>

For qualitative survey items, we find notable differences across samples and preferences. The risk-related qualitative question does not meet the 95% CI criterion in any country, while the time preference question does so in China and Iran. The trust question meets the criterion in Iran and Kenya, whereas altruism and positive reciprocity questions meet it only in China. Table B1 provides the full breakdown of these results, offering precise information on the performance of each item by country and preference. We further conduct robustness checks by restricting the sample to students who answered the control questions correctly. Table B2 in the Appendix shows that the results are very similar and despite the smaller sample size the number of significant coefficients is even slightly higher.

In a comparison between countries, considering the size of coefficients, we find that China and Germany appear to be very close in the domains of risk taking, time discounting, trust, altruism, and positive reci-

<sup>4</sup>For example, Kosfeld and Sharafi (2023) found that the qualitative item PR9 was not significantly correlated with the behavioral measure of positive reciprocity. Here, based on an almost four times larger participant sample, the coefficient is not only slightly larger (0.16 v. 0.13) but also significant on the five percent level.

<sup>5</sup>Like Camerer et al. (2016), we also consider cases in which the entire 95% CI exceeds the original coefficient as successful replications. This is true for the quantitative questions for risk and time in the specification of all replication countries combined without country fixed effects.



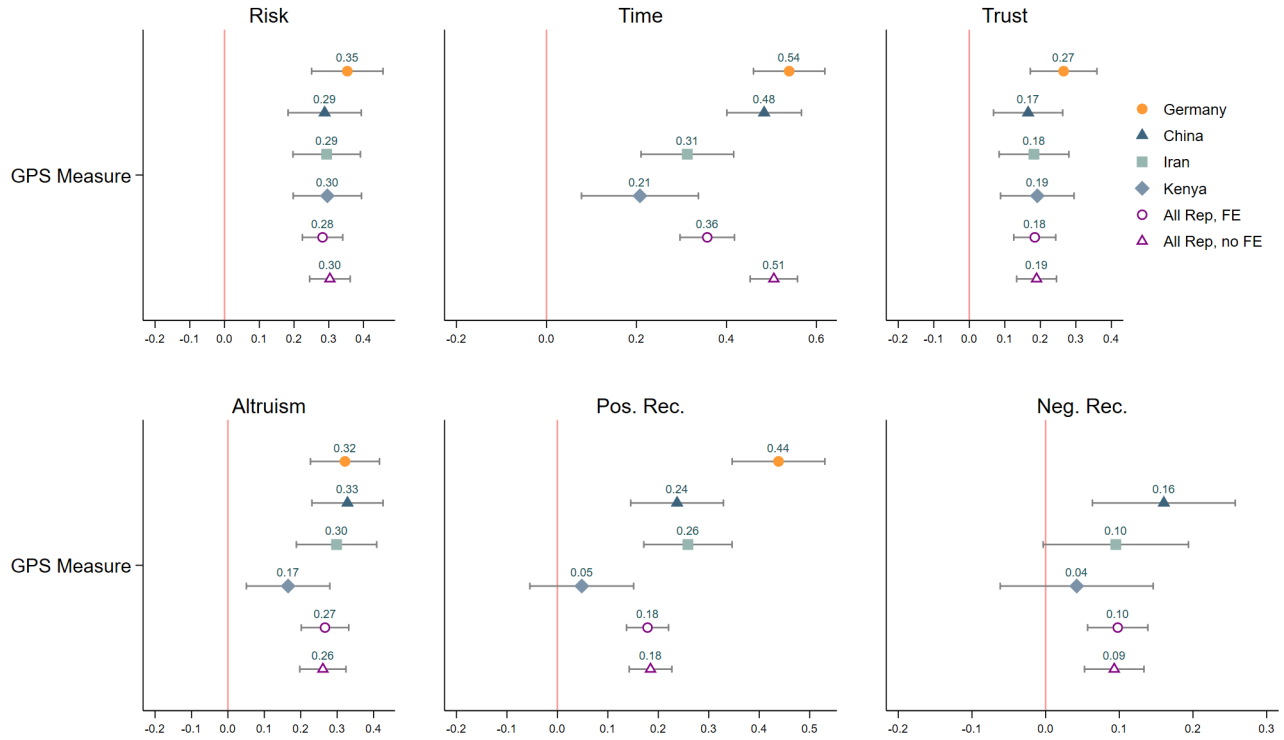


Figure 2: **Plot of regression coefficients of GPS measures.** Coefficients from OLS regressions with 95% confidence intervals. GPS measures are weighted measures combining both GPS items, with weights determined by the corresponding OLS regressions of behavior in the experiments on responses to these survey items in Germany (Falk et al., 2023). For negative reciprocity, the original German sample does not have GPS items NR1 and NR2, therefore, we cannot construct the GPS measure for Germany.

procity. As for negative reciprocity, we are unable to draw inferences because the corresponding GPS items were modified after the original study.

## GPS Measure

Based on the GPS items, Falk et al. (2018) calculate the GPS measure for each preference, which is a weighted measure combining both quantitative and qualitative items, with weights determined by the corresponding OLS regressions from their prior validation study (Falk et al., 2023). Following the same methodology, we can construct the GPS measures for each country and all replication countries combined using the original weights (see Appendix D1) and explore to what extent the GPS measures are associated with preference measures elicited from the experiments.

Figure 2 shows the results. As shown in the plot, the weighted GPS measures are highly predictive of the behavioral measures from the incentivized choice experiments, with the exception of negative reciprocity



again and an overall somewhat lower performance in Kenya. Moreover, the variation in coefficient sizes across countries decreases, in particular for risk taking, trust, and altruism. Table B3 in the Appendix reports the complete results of the corresponding OLS regressions.

We again check the robustness of our results by restricting the sample to students who answered control questions correctly. As can be seen in Table B4 in the Appendix, the results are again very similar with slight improvements. Despite reduced sample sizes, the number of significant coefficients stays the same and the number of coefficients meeting the 95% CI criterion even increases from 12 to 14.<sup>6</sup>

### **3.2 Preference Survey Module (PSM)**

Recall that the GPS items are not the same as the survey items of the Preference Survey Module (PSM) from the original validation study by Falk et al. (2023). The PSM items were selected because they explain the observed variation in behavior in the choice experiments the best. Some of these questions were then refined or replaced for the GPS module to adapt the instrument for a global implementation in the field. Therefore, given the performance of the GPS items and GPS measures shown above, one may expect similar or even better results for the PSM items. Figure 3 confirms this expectation, demonstrating that most PSM items are significantly associated with behavior in the lab. For the complete regression results, see Table B5 in the Appendix.

Consistent with our results for the GPS as well as prior research (Bauer et al., 2020; Kosfeld and Sharafi, 2023), quantitative items consistently and significantly predict choices in the incentivized experiments, with some variation across countries. However, the performance of qualitative items varies considerably across different contexts. Overall, the replication in Kenya continues to show the lowest performance, as all qualitative items, except for trust prove to be statistically insignificant. Looking at the coefficient size, Kenya overall is the most distant from Germany. Nonetheless, we also see greater similarity in the magnitude of coefficients of the quantitative question concerning risk and altruism across all countries. Table B6 in the Appendix presents results based on the sample of students who answered control questions correctly, and the results are again very similar.

---

<sup>6</sup>We can use the GPS and the experimental measure to compare different rankings of the four countries based on the GPS measure from Falk et al. (2018) and the GPS measure or the experimental measure with student samples in our study, respectively. See Table A9 in Appendix A.1 for the results.



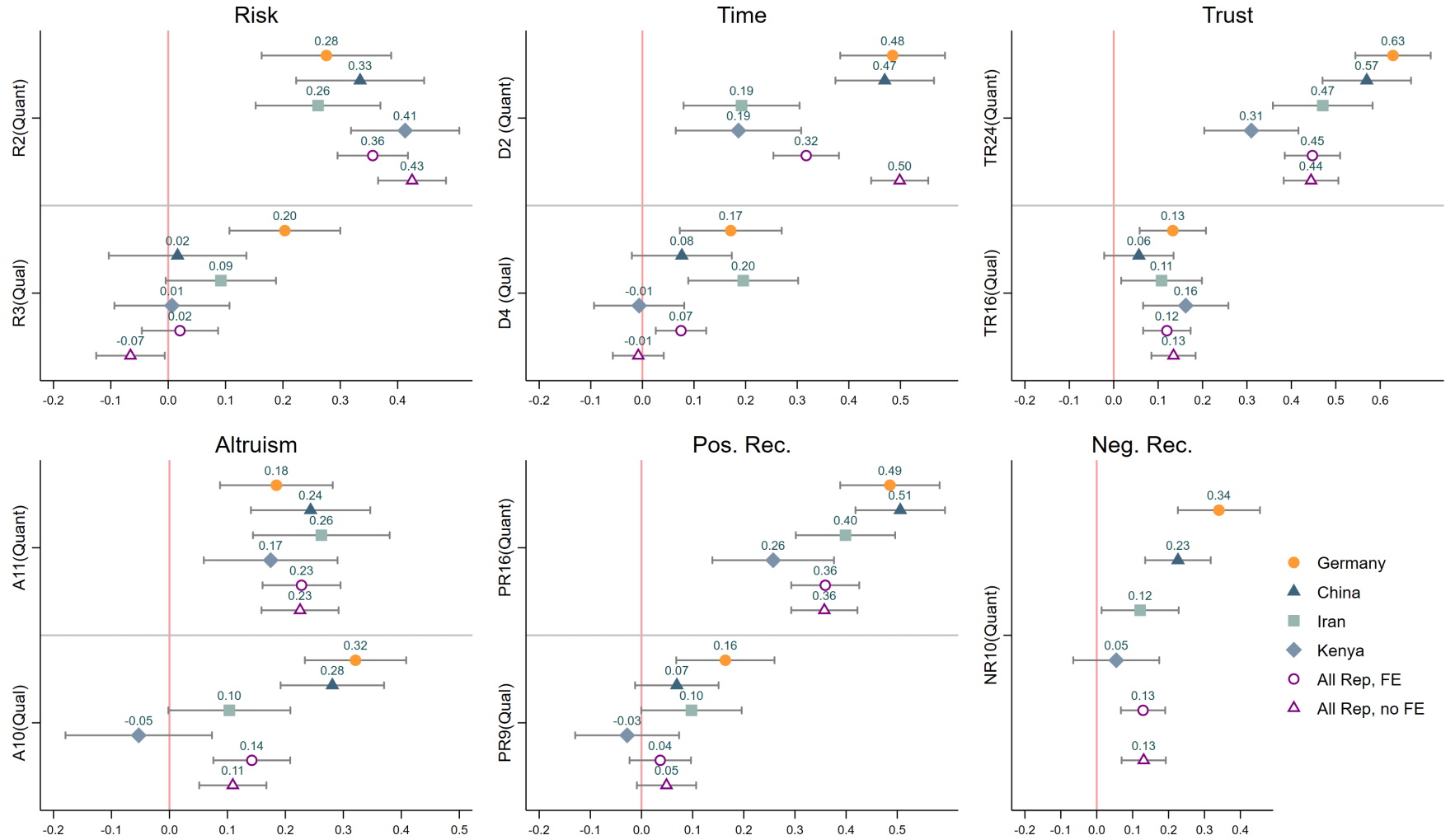


Figure 3: **Plot of regression coefficients of PSM items.** Coefficients from OLS regressions with 95% confidence intervals. For negative reciprocity, the original German sample does not have GPS items NR1 and NR2. R2: Hypothetical multiple price list; R3: willingness to take risks. D2: Hypothetical multiple price list; D4: Willingness to wait in comparison with others. T24: First-mover behavior in trust game; T16: I assume that people have only the best intentions. A11: Hypothetical donation game; A10: Share when it comes to good causes. PR16: Second-mover behavior in trust game; PR9: Thank-you gift in the lost in the city scenario. NR10: Minimum acceptable offer in the hypothetical ultimatum game.



We summarize our first two results as follows.

**Result 1:** *Most GPS and PSM items are significantly associated with behavior in incentivized choice experiments in different countries, except for negative reciprocity in most countries. The performance of quantitative survey items is consistently better than the performance of qualitative survey items.*

**Result 2:** *There exist substantial differences in the magnitude of these associations between other countries and the original study in Germany. Overall, the lowest association for most preferences is found in Kenya.*

### 3.3 Item Selection Based on Linear Combinations

Our next question is whether using the iterative linear item selection procedure deployed in Falk et al. (2023) leads us to identify the same set of questions for predicting corresponding behavior in the lab in a given country. We apply this procedure to each country’s data and present results on the best survey items.

The first step is to run OLS regressions iteratively, regressing the preference elicited in the lab experiment on all possible permutations (up to five items) of the survey items. Using adjusted  $R^2$ , this stage identifies the combinations of survey questions with the highest explanatory power for the choices made in experiments. The second stage assesses the candidate models using information criteria (AIC and BIC) to penalize models with more regressors while minimizing the risk of overfitting. This technique yields the optimal combination of survey items with the highest explanatory power for each preference in each country, with a priority to select a parsimonious model. Thus, by comparing the best combination of survey items identified in each country to the PSM items in the original study, we can examine whether the best combination identified in each country is identical to each other and to that in Germany. Additionally, using the same procedure, we are interested in whether survey items identified with data pooled from all replication countries are identical to those in Germany. Finally, we analyze to what extent survey items selected from data combining *all* four countries (i.e., including Germany) differ from the original PSM items which were only based on the German sample.

Table 2 presents the results sorted by preferences. Here we document, for each preference, the selected combinations of survey items in Germany (which serve as the benchmark), China, Iran, and Kenya, all replication countries combined, and all four countries together. In the last two pooled datasets, we standardize all experimental measures and corresponding survey questions across countries.

Two important observations stand out. First, the quantitative items, consistently emerge as the best can-



didates for explaining behavior in experiments, except for negative reciprocity in Kenya. A possible concern here is that the quantitative items may be always selected because their context and structure resemble those of the incentivized choice experiments, which may indicate a consistency bias. However, consistency bias is less likely to occur when students first answer hypothetical survey questions and then participate in the incentivized choice experiments as decisions in the latter case are payoff relevant. One way to test for consistency bias is to split the sample into two groups, namely, those who participate in the choice experiments during the first session and surveys in the second session (i.e., experiment-first group) and those who complete the survey in the first session and the experiments in the second session (i.e., survey-first group). If consistency bias exists, we would expect the correlation between behavior in the incentivized games and quantitative survey items to be higher in the experiment-first group, compared to the survey-first group. We do not find strong evidence for this. See detailed results in Appendix C.1.

Second, in contrast to the quantitative items, qualitative items vary between samples and there is generally a lack of consensus. For some preferences such as risk and altruism the variation in qualitative items is small. However, for other preferences the variation is large. For example, in the case of negative reciprocity, each sample has its own selected qualitative item.

This leads us to explore an important question: to what extent do various qualitative items selected in different countries differ from each other in terms of their explanatory power for behavior? In other words, how much predictive power do we lose by switching from one qualitative question to another? To answer this question, we pool the data from China, Iran, and Kenya, and plot, for each preference, the sorted adjusted  $R^2$  of all possible combinations of two survey items. As can be seen in Figure 4, very often when the quantitative item is included, we observe a sharp increase in the adjusted  $R^2$ . At the same time, the marginal increase in explanatory power is rather small and often negligible, regardless of which qualitative item is selected in combination with the quantitative item. Put differently, as long as the quantitative item is selected, combining it with any qualitative item does not significantly affect the predictive ability of the combination for the incentivized behavior. In Appendix C.3, we also present the sorted adjusted  $R^2$  of all possible combinations of two items for each preference by country. The overall pattern is very similar. It is worth noting that the maximum adjusted  $R^2$  varies between preferences, ranging from approximately 0.4 for time preferences and negative reciprocity to 0.15 for altruism.

Our third main finding is as follows:

**Result 3:** *Deploying the same item selection procedure as in the original study, we find that the quan-*



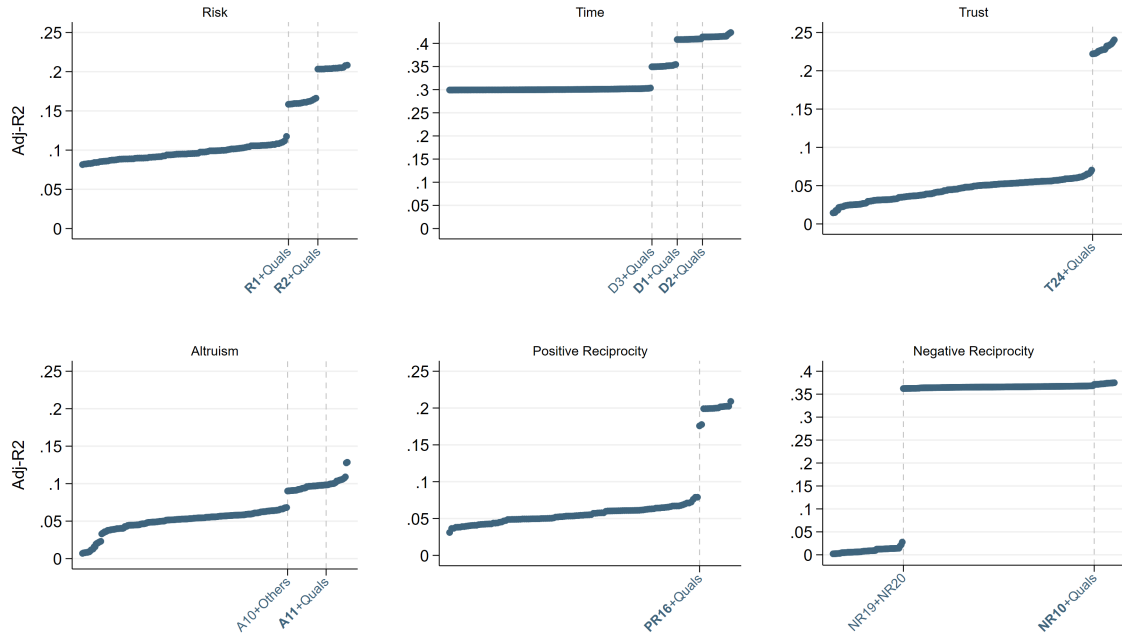


Figure 4: **Sorted adjusted  $R^2$** . This table presents the adjusted  $R^2$ s from OLS regressions of the incentivized measure on different combinations of two-items in the pooled dataset. The bold-faced items represent quantitative questions.

*titative items are consistently selected in each country but the qualitative items vary between countries. However, the overall contribution of qualitative items to explaining behavior in the experiments is relatively low.*

Considering the limitations of the iterative item selection procedure, we next explore alternative approaches and compare their performance in terms of parsimony, effectiveness, and adaptability to various cultures and research needs.



Preference	Items	Countries	Summary
Risk	Hypothetical multiple price lists (R2)	DE, CN, IR, KE, All_Rep, All	R2
	<a href="#">Willingness to take risks (R3)</a>	DE	R3
	Imagine you win 100.000 Euro in a lottery.	CN, IR, All_Rep, All	R16
	You can participate in the lottery using the whole amount you won or only a part of it. (R16)		
Time	<a href="#">Staircase measure (D1)</a>	IR, KE, All_Rep	D1
	Hypothetical multiple price lists (D2)	DE, CN, All_Rep, All	D2
	<a href="#">Willingness to wait? (D3)</a>	IR, All	D3
	Willingness to wait in comparison with others? (D4)	DE	D4
	I am a person who often does not keep their own good resolutions. (D18)	CN	D18
	How much money do you save per month? (D19)	KE	D19
Trust	Hypothetical first-mover behavior (T24)	DE, CN, IR, KE, All_Rep, All	T24
	Most people would take advantage of you when they have the chance (T20)	IR	T20
	Willingness to trust people in your neighborhood (T8)	CN	T8
	Willingness to trust people in your professional environment (T6)	All_Rep	T6
	<a href="#">People have only the best intentions. (T16)</a>	DE, KE, All	T16
Altruism	Hypothetical donation game (A11)	DE, CN, IR, KE, All_Rep, All	A11
	I am willing to donate time and money to charity, even if I do not profit from that directly. (A13)	CN, IR, All_Rep; All	A13
	I am a person who would give their shirt off their back to help others. (A18)	KE	A18
	<a href="#">Share when it comes to good causes. (A10)</a>	DE	A10
Pos. Rec.	Hypothetical second-mover behavior (PR16)	DE, CN, IR, KE, All_Rep, All	PR16
	Willingness to return a favor when it comes to strangers (PR7)	CN, All_Rep; All	PR7
	I do not like the feeling of owning something to someone (PR14)	IR	PR14
	<a href="#">Thank-you gift in the lost in the city scenario (PR9)</a>	DE	PR9
Neg. Rec.	Minimum acceptable offer in ultimatum game (NR10)	DE, CN, IR, KE, All_Rep, All	NR10
	Willingness to punish unfair behavior when it comes to your professional environment. (NR7)	IR	NR7
	If someone harms me on purpose, I will try to give that person a taste of their own medicine (NR14)	CN	NR14
	Willing to punish unfair behavior even if this is costly	DE	NR1/2
	Willingness to punish unfair behavior when it comes to your neighborhood. (NR9)	KE	NR9
	Willingness to punish your circle of friends (NR6)	KE	NR6
	When someone puts me into a difficult situation, I will do the same to them (NR12)	All	NR12
	Willingness to punish someone who treats you unfairly (NR1)	All_Rep	NR1
Number of observations: DE: 382(360); CN: 395; IR: 389; KE: 383, All (replication combined): 1167, All: 1549 (1527)			

Table 2: **Best combination of survey items.** This table shows the best combination of survey items based on the item selection procedure in Falk et al. (2023). Blue-colored items are GPS items.



## 4 Developing Alternative Survey Modules

While the iterative linear item selection procedure selects the combination of two items that has the highest predictive power, an important caveat of this method is that it ignores whether or not the marginal increase in adjusted  $R^2$  between different models is meaningful. In this section, we employ other approaches to develop alternative survey modules and better understand how various modules perform in terms of their parsimony, reliability in adapting to different cultures and research needs, and effectiveness in explaining behavior in choice experiments. Again, we utilize data from China, Iran, Kenya, and Germany separately, as well as data from all countries combined.

For each preference, we present alternative modules derived from two types of analysis. First, we report a single-item module. This module consists of individual quantitative or qualitative survey items that are most effective in predicting experimental choices compared to other survey questions. We run univariate OLS regressions of the standardized experimental measure on each corresponding standardized item for every country. Instead of using the highest adjusted  $R^2$  as the selection criterion, we select candidate items that are significantly associated with behavior at least at a five-percent level *in all four countries*. If no item meets this criterion, we shortlist items that are significantly associated with behavior in three out of the four countries. In this way, we ensure that the candidate items selected for each preference have high reliability across different cultures and contexts. After identifying candidates that meet this criterion, we regress the experimental measure on each candidate with country fixed effects using the pooled dataset and report the three questions with the highest adjusted  $R^2$ , again being significant at the five-percent level.<sup>7</sup> Given the strong performance of quantitative items, they are always included in the reported questions for all preferences. However, we also identify qualitative items. For negative reciprocity, only two questions pass our inclusion criteria.

Second, we develop a multiple-item module derived from the application of the Least Absolute Shrinkage and Selection Operator (Lasso) technique. Feature selection involves choosing a reduced number of explanatory variables (i.e., a subset of relevant features) to describe a response variable (Guyon and Elisseeff, 2008). This method helps eliminate redundant or irrelevant features, as well as those strongly correlated in the data while preserving valuable information. Our selection process employs cross-validation (CV) to identify the most parsimonious model that minimizes out-of-sample prediction error. For this analysis, we

---

<sup>7</sup>In the pooled dataset, we again standardize all experimental measures and corresponding survey questions across countries.



set the penalty ( $\lambda$ ) to a level that consistently yields a module with exactly three items, thereby maintaining parsimony in the model. For risk and time preferences, we identify two quantitative items—multiple price lists and staircase. As the correlation between these two items exceeds 0.7 in our dataset, we exclude the longer measure and include only the staircase questions.

To evaluate the relative performance of both modules, we use two metrics. First, we report their explanatory power (i.e., adjusted  $R^2$ ) for the choices in the experiments using the combined data from all countries. Second, to indicate how well the model generalizes to new, unseen data, we randomly split our combined data into training (75%) and test (25%) sets, and compare each module’s out-of-sample prediction accuracy as measured by the Mean Squared Error (MSE). The lower the MSE, the more precise the prediction is on new unseen observations.

Table 3 presents both alternative survey modules for the measurement of risk preferences, time discounting, trust, altruism, as well as positive and negative reciprocity. In the single-item module, the quantitative items consistently show the highest performance in predicting behavior with a noticeable gap to the next best. This suggests that the hypothetical versions can be used as good proxies for the incentivized experiments. Other items hardly come close to matching the predictive power of simply using the hypothetical versions of the games. Even among the quantitative single-item modules, some have advantages over others. For instance, in measuring time preference, the best single quantitative item, namely, the multiple price list (D2), has higher predictive power than the next single quantitative item, the staircase question (D1). However, the latter is obviously more parsimonious requiring only five decisions compared to 31 decisions in D2. We suggest researchers make a situation-specific decision when it comes to which items to use, balancing predictive power with parsimony.

Table 3 shows that there are quite a number of GPS items (highlighted in blue) selected for both modules. This indicates that the GPS items are generally good candidates for measuring preferences, at least compared to the other questions in the batteries. While we cannot rule out the possibility that there might be other survey items outside the batteries that perform better than GPS, the GPS items are at least effective within the scope of those put together by Falk et al. (2023) and included in our analysis.

Finally, there are overlaps between the different modules, each with its own strengths and weaknesses regarding parsimony and accuracy. The multiple-item modules identified by Lasso consistently have the lowest MSE, indicating the highest accuracy in out-of-sample prediction—an important criterion in the context of new and unseen data. However, this approach is less parsimonious than the single-item module.



Preference	Module	Item	Countries	Adj. $R^2$	MSE test
Panel A: Risk	Single-item	R2: Hypothetical multiple price lists (Quant)	CN; DE; IR; KE	0.203	.895
		<a href="#">R1: Switching row from staircase measure (Quant)</a>	CN; DE; IR; KE	0.159	.951
		R16: you win 100,000 EUR in a lottery. How much would you invest in the risky yet profitable lottery? (Qual)	CN; DE; IR; KE	0.106	1.002
	Multiple-item (Lasso)	<a href="#">R1: Switching row from staircase measure</a> <a href="#">R3: Willingness to take risks</a> R15: How well do the following statements describe you as a person? I like taking risks		0.166	.944
Panel B: Time	Single-item	D2: Hypothetical multiple price lists (Quant)	CN; DE; IR; KE	0.413	0.596
		<a href="#">D1: Switching row from staircase measure (Quant)</a>	CN; DE; IR; KE	0.408	0.610
		<a href="#">D3: Willingness to wait (Qual)</a>	CN; DE; IR; KE	0.350	0.671
	Multiple-item (Lasso)	<a href="#">D1: Switching row from staircase procedure</a> <a href="#">D3: Willingness to wait</a> D9: willingness to wait wrt. bigger purchases		0.420	0.593
Panel C: Trust	Single-item	T24: First-mover behavior in trust game (Quant)	CN; DE; IR; KE	0.222	0.769
		T8: Willingness to trust neighbours (Qual)	CN; DE; IR; KE	0.054	0.967
		<a href="#">T16: I assume that people have only the best intentions (Qual)</a>	CN; DE; IR; KE	0.054	0.960
	Multiple-item (Lasso)	T6: Willingness to trust people in your professional environment <a href="#">T16: I assume that people have only the best intentions</a> T24: 1st mover decision in hypothetical trust game		0.246	0.750

Table 3: **Alternative Preference Modules.** This table reports alternative survey modules for each preference. Items highlighted in blue are GPS survey items. Country codes show in which countries items selected in each module are significantly correlated with experimental choices. The adjusted  $R^2$  is from the data with all four countries combined. The Mean Squared Error (MSE) shows the precision of the out-of-sample prediction. The lower the number, the more precise the prediction is on unseen data.



Panel D: Altruism	Single-item	<a href="#">A11: Hypothetical donation game (Quant)</a>	CN; DE; IR; KE	0.096	0.866
		A13: I am willing to donate time and money to charity, even if I do not profit from that directly (Qual)	CN; DE; IR	0.096	0.890
		<a href="#">A10: Willingness to give to good causes (Qual)</a>	CN; DE; IR	0.090	0.881
	Multiple-item (Lasso)	<a href="#">A10: Willingness to give to good causes</a> <a href="#">A11: Hypothetical donation game</a> A13: I am willing to donate time and money to charity, even if I do not profit from that directly		0.134	0.844
Panel E: Pos. Reciprocity	Single-item	Quant - PR16: 2nd mover behavior in hypothetical (Quant) trust game	CN; DE; IR; KE	0.198	0.852
		<a href="#">PR9: Thank-you gift in the lost city scenario (Qual)</a>	CN; DE; IR	0.062	1.018
		<a href="#">PR11: When someone does me a favor (Qual)</a>	CN; DE; IR	0.060	1.005
	Multiple-item (Lasso)	PR7: Willingness to return a favor to strangers <a href="#">PR9: Thank-you gift in the lost in the city scenario</a> PR16: 2nd mover behavior in hypothetical trust game		0.208	0.853
Panel F: Neg. Reciprocity	Single-item	NR10: Minimum acceptable offer in hypothetical Ultimatum Game (Quant)	CN; DE; IR	0.371	0.682
		NR12: When someone puts me into a difficult situation, I will do the same to them (Qual)	CN; DE; IR	0.367	0.694
	Multiple-item (Lasso)	NR10: Minimum acceptable offer in hypothetical Ultimatum Game <a href="#">NR11: Take revenge at the first occasion</a> NR12: When someone puts me into a difficult situation, I will do the same to them		0.374	0.694

Table 3: **Alternative Preference Module - continued.**



## 5 Exploring the Interactions of Different Measures Using PCA

To better understand the interplay between incentivized measures and quantitative and qualitative survey questions, we combine all available data for each preference—survey responses and experimental measures—in this section. We apply principal component analysis (PCA) to identify components that highlight distinct clusters of highly correlated items. This approach enables us to explore the dataset’s underlying structure, reduce its dimensionality, and pinpoint the primary components driving each preference. PCA is a widely used linear dimensionality reduction technique that transforms original variables into a new set of components. These components are constructed as linear combinations of the original variables, designed to capture the maximum possible variance inherent in the data and the underlying variable construct—in our case, preferences.

In line with established practice in the literature, Table 4 documents all components with eigenvalues greater than one (Kaiser, 1960), presented in order of the proportion of variance they explain (Abdi and Williams, 2010). To facilitate interpretation, we apply a varimax rotation to the resulting components (Furr, 2021). In Appendix C.4, we list the questions corresponding to each component by preference. The first principal component primarily aggregates clusters of survey questions related to individuals’ willingness to take risks, wait, trust others, share with others, return a favor, and punish those who treat others unfairly. These dimensions correspond to measures of risk-taking, time discounting, trust, altruism, and both positive and negative reciprocity. Each of these dimensions is typically constructed from various self-assessment survey items of the respective preference across different domains and contexts.

This analytical approach maximizes the use of available data to unveil the underlying structure and latent factors associated with each preference. On average, we find that the resulting modules consist of approximately four overarching components per preference. Since the share of variance explained by each component is endogenous to the survey’s construction (e.g., the number of questions with similar context and content), we focus less on the order of the components and more on how the questions cluster together.

One important finding from this analysis is that for all preferences the quantitative measures—including both the incentivized experimental measure and the hypothetical survey version—bundle together exclusively. Only for time preferences, altruism, and positive reciprocity do we see an exception, where they are combined with a qualitative survey item, though with relatively small loadings. Since the identified components are orthogonal, this suggests that quantitative and qualitative measures contribute distinct in-



Preferences	Principal components	Fraction of Explained Variance
A: Risk taking	Willingness to take risks	0.327
	Quantitative measures for risk	0.121
	Risk in daily activities	0.089
	Risk in financial investments	0.087
B: Time discounting	Willingness to wait regarding financial decision	0.160
	Quantitative measures for patience	0.119
	Willingness to wait in general	0.112
	Tendency to postpone tasks	0.104
	Compensation for giving up a vacation trip	0.053
C: Trust	Willingness to trust	0.260
	Most people are fair	0.085
	Building trust in other people and their intentions	0.078
	Most people act in their own best interest	0.069
	Quantitative measure for trust	0.060
	Willingness to rely on strangers	0.050
	Willingness to lend to strangers	0.040
D: Altruism	Willingness to share with people	0.354
	Willingness to do good if I expect it returns	0.097
	Quantitative measure for altruism	0.060
	Volunteer to act for good causes	0.060
	I don't understand why some people fight for a cause without direct benefits	0.054
E: Positive reciprocity	Willingness to return a favor	0.347
	I help someone who has helped me before	0.187
	Thank-you gift to return a favor	0.100
	Quantitative measures for positive reciprocity	0.081
F: Negative reciprocity	Willingness to punish people with unfair behavior	0.225
	Willingness to punish people who treated others unfairly	0.224
	Dislike being taken for a fool and pushed around	0.107
	Quantitative measures for negative reciprocity	0.049
	Fairness in hypothetical scenario	0.045

Table 4: **Principal components for each preference.** This table presents the results of principal component analysis. For each preference, we present components with eigenvalue larger than one and present them in the order of the size of their fraction of explained variance. In Appendix C.4, we list the individual questions corresponding to each component by preference.



formation to the data structure, with each type of measure capturing different aspects of the latent construct, i.e., preference.

## 6 Predicting Real-Life Behavior with Different Preference Measures

Research has documented the predictive power of economic preferences for human decision-making and its associated life outcomes across various social and economic domains (cf. references in the introduction of this paper). In this section, we return to this point and use our data to answer the question which type of measurement, behavioral or survey-based, performs better in such prediction tasks.

For example, prior literature has shown that women tend to be more risk averse and more altruistic than men (Byrnes et al., 1999; Croson and Gneezy, 2009). By regressing the experimental measure as well as the different single-item modules from Section 4 on gender, we broadly confirm the gender difference in risk taking regardless of what preference measure we use (see Table 5, upper panel). However, we only find supportive evidence for a significant gender difference in altruism when the latter is measured by means of survey questions (see Table 5, lower panel).<sup>8</sup>

Next, using the data from our survey on real-life behavior we analyze the association between preferences and typical behaviors that can be hypothesized to be driven by underlying preference heterogeneity. Specifically, we consider risky behavior in the domains of health and finance, such as smoking and investment in securities, that have been shown to correlate with individuals' willingness to take risks (Anderson and Mellor, 2008; Dohmen et al., 2011). We also consider the association between altruism and volunteering as well as pro-environmental behavior (Cárdenas et al., 2013; Lades et al., 2021), the association between patience and saving behavior (Sutter et al., 2013), and the association between social capital as measured by the number of social connections and trust (Dohmen et al., 2009).

Table 6 reports the results. Again, we compare the behavioral measures based on incentivized choice experiments as well as the different single-item modules from Section 4 for each preference and behavior. For risk preferences, we additionally include the qualitative item R3 (willingness to take risks).

The real-life outcomes used in this analysis are binary or standardized variables. Both the survey and experimental variables are standardized, consistent with the other analyses throughout the paper. The regressions are conducted on the pooled dataset from three countries, incorporating country fixed effects.

---

<sup>8</sup>To address possible measurement error in the experimental measures, we restrict the sample to students who answered all control questions correctly in the following.



Risk					
	Experiment	R1: Staircase measure	R2: Hypothetical multiple price lists	R3: Willingness to take risks	R16: you win 100,000 EUR in a lottery. . .
Female	-0.191*** (0.060)	-0.147** (0.062)	-0.236*** (0.059)	-0.349*** (0.056)	-0.289*** (0.061)
Adj $R^2$	0.086	0.037	0.068	0.180	0.074
N	984	984	984	984	984
Altruism					
	Experiment	A11: Hypothetical donation game	A10: Willingness to give to good causes	A13: I am willing to donate time and money...	
Female	-0.001 (0.063)	0.035 (0.063)	0.227*** (0.060)	0.273*** (0.059)	
Adj $R^2$	0.001	0.063	0.147	0.158	
N	1003	1003	1003	1003	

Table 5: **Regression coefficients of gender on preferences.** In this table, we document OLS regression coefficients of gender on various measures of risk and altruism. The data is the pooled sample of China, Iran and Kenya and only contains participants who answered control questions correctly. We controlled for country fixed effects. Items highlighted in blue are GPS survey items. \*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$ .

Two observations stand out. First, survey-based preference measures are consistently and significantly associated with real-life behaviors in the domains of smoking, financial investment, volunteering, pro-environmental behavior, savings, and social connections. Second, preferences elicited from incentivized choice experiments, on the other hand, are uncorrelated or only weakly associated with real-life behaviors except for savings. In general, it seems that qualitative survey items do somewhat better than both the quantitative items and the behavioral measure, but there is no clear overall picture. For example, for risk and time preferences, quantitative items do rather well. Further, it seems plausible that qualitative items which are based on self-assessments, have an advantage in this horse race, because they are elicited with the same method as the self-reports on participants' real-life behavior. Therefore, the results should be interpreted with some care.

We summarize our last finding as follows:

**Result 4:** *Overall, survey-based preference measures seem to be better predictors for (self-reported) real-life behaviors than preference measures based on incentivized choice experiments.*



<b>Risk (N = 984)</b>		<b>Smoking</b>	<b>Investment</b>
Experiment		0.003 (0.009)	0.013 (0.016)
Adjusted R-squared		0.100	0.018
<b>R1: Staircase measure</b>		0.033*** (0.010)	0.001 (0.015)
Adjusted R-squared		0.111	0.017
R2: Hypothetical multiple price lists		0.034*** (0.009)	0.021 (0.015)
Adjusted R-squared		0.111	0.019
<b>R3: Willingness to take risks</b>		0.026*** (0.010)	0.072*** (0.016)
Adjusted R-squared		0.106	0.037
R16: you win 100,000 EUR in a lottery. . .		0.012 (0.009)	0.042*** (0.016)
Adjusted R-squared		0.101	0.025
<b>Altruism (N = 1003)</b>		<b>Social contribution</b>	<b>Pro-environment</b>
Experiment		0.002 (0.013)	-0.004 (0.024)
Adjusted R-squared		0.292	0.418
<b>A11: Hypothetical donation game</b>		0.011 (0.012)	0.068*** (0.022)
Adjusted R-squared		0.293	0.422
<b>A10: Willingness to give to good causes</b>		0.030** (0.014)	0.046* (0.027)
Adjusted R-squared		0.296	0.419
A13: I am willing to donate time and money...		0.063*** (0.014)	0.068** (0.028)
Adjusted R-squared		0.306	0.421
<b>Time (N = 1121)</b>		<b>Saving</b>	
Experiment		0.048*** (0.015)	
Adjusted R-squared		0.047	
<b>D1: Staircase measure</b>		0.034** (0.014)	
Adjusted R-squared		0.042	
D2: Hypothetical multiple price lists		0.042*** (0.014)	
Adjusted R-squared		0.044	
<b>D3: Willingness to wait</b>		0.048*** (0.013)	
Adjusted R-squared		0.048	
<b>Trust (N = 992)</b>		<b>Social connection</b>	
Experiment		0.035 (0.031)	
Adjusted R-squared		0.032	
T24: 1st mover in a hypothetical trust game		0.043 (0.032)	
Adjusted R-squared		0.033	
<b>T16: People have only the best intentions</b>		0.076** (0.033)	
Adjusted R-squared		0.037	
T8: Willingness to trust neighbors		0.150*** (0.035)	
Adjusted R-squared		0.052	

Table 6: **Regression coefficients of preferences on behaviors.** In this table, we report OLS regressions of various preference measures on real-life behavior. The data is the pooled sample of China, Iran and Kenya and contains only participants who answered control questions correctly. Items highlighted in blue are GPS survey items. We controlled for country fixed effects. \*\*\*  $p < 0.01$ ; \*\*  $p < 0.05$ ; \*  $p < 0.1$ .



## 7 Discussion and Conclusion

Economists traditionally utilize incentivized choice experiments as a benchmark for measuring economic preferences, following the revealed-preference approach. The assumption is that by choosing between available options, individuals effectively reveal their preference over these options. While such choice-based, i.e., behavioral, measures closely connect to economic theory and sometimes also particular model predictions, they come with their own challenges. For example, incentivized behavioral measures are costly and difficult to implement on a large scale. Additionally, their instructions are sometimes hard to comprehend, even for students at top universities. Furthermore, it may be uncertain to what extent researchers are able to control the various mental representations participants form when facing relatively abstract experimental scenarios, nor how these perceptions influence their decision making (Detemple, 2024). These uncertainties raise concerns about the extent to which behavioral measures are capable of accurately and thoroughly capturing the true construct of economic preferences.

Survey instruments, on the other hand, also pose their own challenges. For instance, responses to survey questions may be prone to reference bias (Heckman and Kautz, 2021; Kautz et al., 2014; Lira et al., 2022) or self-serving biases (Gino et al., 2016). In addition, inattention and strategic motives can cause imprecise responses (Dohmen et al., 2011). Survey questions are also not incentive-compatible by design, which makes economists particularly skeptical of their reliability. However, they are highly portable tools that can be quickly deployed in different field settings and often be better adapted to specific contexts and circumstances than behavioral experiments.

In this study, we explore the relationship between experimental measures and survey items by constructing a unique dataset that includes incentivized experimental measures, quantitative survey questions, and qualitative self-assessment surveys across a range of economic preferences. This dataset spans multiple countries and is complemented by participants' demographic information as well as data on their real-life behaviors. Leveraging this comprehensive dataset, we present several new insights and results that contribute to the literature on preference measurement.

Our replication of the studies by Falk et al. (2018, 2023) shows that GPS and PSM items generally predict behavior in incentivized choice experiments across a variety of countries beyond Germany, where the original validation was conducted, with exceptions for negative reciprocity in most countries and almost all preferences in Kenya. Quantitative items consistently outperform qualitative items in predictive strength,



though we generally observe notable variations in the strength of associations across countries. It is worth mentioning that most coefficients in these associations fall below 0.4, raising an important question: Do these magnitudes suffice for qualifying these items as effective measures of economic preferences? Our findings suggest that there is room for improvement, emphasizing the potential need for stronger alternatives or refinements in preference measurement tools.

In replicating the iterative linear item selection procedure from Falk et al. (2023), we find that while the quantitative item is consistently selected across all countries, the specific qualitative item varies. Yet, once the quantitative item is included as a predictor different qualitative items often add only minimal explanatory power on top. This suggests that the quantitative item alone carries critical predictive weight, with qualitative items providing limited additional value. The consistency of the quantitative item across diverse contexts thus underscores its robustness as a reliable predictor.

In addition to this replication exercise, we develop alternative survey modules that encompass both single-item and multiple-item formats. Given that these modules are designed to be utilized in future research across various contexts, we focus not only on their predictive power for incentivized behavior but also prioritize their reliability, as demonstrated by consistent associations with behavior across *all* countries in our sample. This dual focus equips researchers with a hopefully practical tool for selecting the most suitable module based on specific research needs and contexts.

To study in more detail the interplay between experimental and survey-based preference elicitation, we conduct an exploratory principal component analysis (PCA), through which we extract core and orthogonal components that capture the latent constructs of each preference. This analysis reveals that quantitative and qualitative measures are always clustered in separate components, suggesting they capture different aspects of the same preference.

Finally, we find that survey-based preference measures are consistently associated with real-life behaviors across several domains, while experimental measures show only weak correlations with real-life behaviors in this study. We discuss each side of this finding in more detail. One explanation for the weaker correlation between experimental choices and real-life behavior may be that we lack sufficient control over how individuals perceive and interpret the study context, in particular when instructions emphasize from the outset that their choices will affect their payoffs. This awareness may stimulate participants to adopt a more payoff-maximizing approach, engaging in analytical or “slow” thinking characterized by deliberate and reflective decision-making. In contrast, people often rely on more automatic, “fast” thinking in daily life,



where decisions are typically less calculated and more compatible with immediate intuitions (Kahneman, 2011). This distinction raises an important question: Which preferences should we prioritize measuring, the ones that individuals reveal through deliberate calculated decision-making, or the preferences underlying their everyday actions and decisions?

The other side of this finding is the meaningful correlation of real-life outcomes and survey questions. One concern here is that in our study real-life economic outcomes are self-reported, and since the elicitation method closely resembles survey measurement, we should be cautious in over-interpreting these results. However, there exist several studies that use observational data to indicate real-life behavior and relate it to both experimental and survey measures and found survey measure to outperform the experimental measures (Dohmen et al., 2011; Coppola, 2014; Hoffmann et al., 2015; Verschoor et al., 2016; Menkhoff and Sakha, 2017; Pinger, 2017; Charness et al., 2020). In addition, the results in Falk et al. (2018) may also help mitigate this concern. Their research validates the employed survey measures by showing significant correlations between preference measures and various *country-level* economic outcomes derived from observational data on a global scale. This suggests that the survey items are capable of correctly capturing preference variations. However, Falk et al. (2018) only report correlations for the combined measure, leaving some ambiguity as to whether these correlations are primarily driven by the combination of quantitative and qualitative questions or if one type plays a dominant role. Assessing the performance of qualitative self-assessment questions and quantitative items separately could provide valuable insights, clarifying how effectively each type captures variation in these economic outcomes.

We aim for our current study to serve as an initial investigation into which measurement method is the most precise and effective at capturing the latent construct of economic preferences—behavioral experiments, surveys, or a combination of both? We hope our work will advance the research agenda in this field in a constructive and productive way. Currently, we are incorporating data from two additional countries, Colombia and the United States, to enhance the diversity of our sample, enrich our analysis, and achieve more comprehensive global coverage. We will also establish a platform to make our study materials accessible to other researchers in the future. This platform aims to lay the groundwork for others to join and provide a foundational framework for a comprehensive validation and development of effective measurement tools. Additionally, it may facilitate further research into how different cultures vary in specific economic preferences and the possible origins of these differences.



## References

- ABDI, H. AND L. J. WILLIAMS (2010): “Principal component analysis,” *WIREs Computational Statistics*, 2, 433–459.
- ANDERSON, L. R. AND J. M. MELLOR (2008): “Predicting health behaviors with an experimental measure of risk preference,” *Journal of Health Economics*, 27, 1260–1274.
- BACKES-GELLNER, U., H. HERZ, M. KOSFELD, AND Y. OSWALD (2021): “Do preferences and biases predict life outcomes? Evidence from education and labor market entry decisions,” *European Economic Review*, 134, 103709.
- 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, 103493.
- BERG, J., J. DICKHAUT, AND K. MCCABE (1995): “Trust, Reciprocity, and Social History,” *Games and economic behavior*, 10, 122–142.
- BONIN, H., T. DOHMEN, A. FALK, D. HUFFMAN, AND U. SUNDE (2007): “Cross-sectional earnings risk and occupational sorting: The role of risk attitudes,” *Labour Economics*, 14, 926–937.
- BRISLIN, R. W. (1970): “Back-translation for cross-cultural research,” *Journal of Cross-cultural Psychology*, 1, 185–216.
- BYRNES, J. P., D. C. MILLER, AND W. D. SCHAFER (1999): “Gender differences in risk taking: A meta-analysis,” *Psychological Bulletin*, 125, 367–383.
- CALIENDO, M., F. M. FOSSEN, AND A. S. KRITIKOS (2009): “Risk attitudes of nascent entrepreneurs—new evidence from an experimentally validated survey,” *Small business economics*, 32, 153–167.
- CAMERER, C. F., A. DREBER, E. FORSELL, T.-H. HO, J. HUBER, M. JOHANNESSON, M. KIRCHLER, J. ALMENBERG, A. ALTMEJD, T. CHAN, AND E. HEIKENSTEN (2016): “Evaluating replicability of laboratory experiments in economics,” *Science*, 351, 1433–1436.
- CÁRDENAS, J. C., A. CHONG, AND H. ÑOPO (2013): “Stated social behavior and revealed actions: Evidence from six Latin American countries,” *Journal of Development Economics*, 104, 16–33.
- CARPENTER, J. AND E. SEKI (2011): “Do social preferences increase productivity? Field experimental evidence from fishermen in Toyama Bay,” *Economic Inquiry*, 49, 612–630.
- CHABRIS, C. F., D. LAIBSON, C. L. MORRIS, J. P. SCHULDT, AND D. TAUBINSKY (2008): “Individual laboratory-measured discount rates predict field behavior,” *Journal of risk and uncertainty*, 37, 237–269.
- CHAPMAN, J., P. ORTOLEVA, E. SNOWBERG, L. YARIV, AND C. CAMERER (2024): “Reassessing Experimental Validation and Qualitative Self-Assessments,” Working paper.
- CHARNESS, G., T. GARCIA, T. OFFERMAN, AND M. C. VILLEVAL (2020): “Do measures of risk attitude in the laboratory predict behavior under risk in and outside of the laboratory?” *Journal of Risk and Uncertainty*, 60, 99–123.



- CHEN, D. L., M. SCHONGER, AND C. WICKENS (2016): “oTree—An open-source platform for laboratory, online, and field experiments,” *Journal of Behavioral and Experimental Finance*, 9, 88–97.
- COPPOLA, M. (2014): “Eliciting risk-preferences in socio-economic surveys: How do different measures perform?” *The Journal of Socio-Economics*, 48, 1–10.
- CROSON, R. AND U. GNEEZY (2009): “Gender differences in preferences,” *Journal of Economic Literature*, 47, 448–474.
- DETEMPLE, J. (2024): “Thoughts about the dictator and trust game,” *SAFE Working Paper No. 422*.
- DOHMEN, T., A. FALK, D. HUFFMAN, AND U. SUNDE (2009): “Survey evidence on behavioural outcomes,” *The Economic Journal*, 119, 592–612.
- DOHMEN, T., A. FALK, D. HUFFMAN, U. SUNDE, J. SCHUPP, AND G. G. WAGNER (2011): “Individual risk attitudes: Measurement, determinants, and behavioral consequences,” *Journal of the European Economic Association*, 9, 522–550.
- FALK, A., A. BECKER, T. DOHMEN, B. ENKE, D. HUFFMAN, AND U. SUNDE (2018): “Global evidence on economic preferences,” *The Quarterly Journal of Economics*, 133, 1645–1692.
- 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, 1935–1950.
- FALK, A., E. FEHR, AND U. FISCHBACHER (2005): “Driving forces behind informal sanctions,” *Econometrica*, 73, 2017–2030.
- FRIEBEL, G., M. KOSFELD, AND G. THIELMANN (2019): “Trust the police? Self-selection of motivated agents into the German police force,” *American Economic Journal: Microeconomics*, 11, 59–78.
- FURR, R. M. (2021): *Psychometrics: An introduction*, SAGE publications.
- GINO, F., M. I. NORTON, AND R. A. WEBER (2016): “Motivated Bayesians: Feeling moral while acting egoistically,” *Journal of Economic Perspectives*, 30, 189–212.
- GOLSTEYN, B. H., H. GRÖNQVIST, AND L. LINDAHL (2014): “Adolescent time preferences predict life-time outcomes,” *The Economic Journal*, 124, 739–761.
- GRUND, C. AND D. SLIWKA (2010): “Evidence on performance pay and risk aversion,” *Economics Letters*, 106, 8–11.
- GÜTH, W., R. SCHMITTBERGER, AND B. SCHWARZE (1982): “An experimental analysis of ultimatum bargaining,” *Journal of Economic Behavior & Organization*, 3, 367–388.
- GUYON, I. AND A. ELISSEEFF (2008): “An introduction to variable and feature selection,” *Journal of Machine Learning Research*, 3, 1157–1182.
- HANUSHEK, E. A., L. KINNE, P. LERGETPORER, AND L. WOESSMANN (2022): “Patience, risk-taking and human capital investment across countries,” *The Economic Journal*, 132, 2290–2307.
- HECKMAN, JAMES J., J. T. AND T. KAUTZ (2021): *Some Contributions of Economics to the Study of Personality*, New York, NY: Guilford Press.



- HOFFMANN, A. O., T. POST, AND J. M. PENNINGS (2015): “How investor perceptions drive actual trading and risk-taking behavior,” *Journal of Behavioral Finance*, 16, 94–103.
- JAEGER, D. A., T. DOHMEN, A. FALK, D. HUFFMAN, U. SUNDE, AND H. BONIN (2010): “Direct evidence on risk attitudes and migration,” *The Review of Economics and Statistics*, 92, 684–689.
- KAHNEMAN, D. (2011): “Thinking, fast and slow,” *Farrar, Straus and Giroux*.
- KAISER, H. F. (1960): “The application of electronic computers to factor analysis,” *Educational and psychological measurement*, 20, 141–151.
- KAUTZ, T., J. J. HECKMAN, R. DIRIS, B. TER WEEL, AND L. BORGHANS (2014): “Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success,” *NBER Working Paper No. 20749*.
- KIRBY, K. N. AND N. M. PETRY (2004): “Heroin and cocaine abusers have higher discount rates for delayed rewards than alcoholics or non-drug-using controls,” *Addiction*, 99, 461–471.
- KOSFELD, M. AND D. RUSTAGI (2015): “Leader punishment and cooperation in groups: Experimental field evidence from commons management in Ethiopia,” *American Economic Review*, 105, 747–783.
- KOSFELD, M. AND Z. SHARAFI (2023): “The Preference Survey Module: evidence on social preferences from Tehran,” *Journal of the Economic Science Association*, 1–13.
- LADES, L. K., K. LAFFAN, AND T. O. WEBER (2021): “Do economic preferences predict pro-environmental behavior?” *Ecological Economics*, 183, 106977.
- LIRA, B., J. M. O’BIEN, P. A. PEÑA, B. M. GALLA, S. D’MELLO, D. S. YEAGER, A. DEFNET, T. KAUTZ, K. MUNKACSY, AND A. L. DUCKWORTH (2022): “Large studies reveal how reference bias limits policy applications of self-report measures,” *Scientific Reports*, 12, 19189.
- MEIER, S. AND C. SPRENGER (2010): “Present-biased preferences and credit card borrowing,” *American Economic Journal: Applied Economics*, 2(1), 193–210.
- MENKHOFF, L. AND S. SAKHA (2017): “Estimating risky behavior with multiple-item risk measures,” *Journal of Economic Psychology*, 59, 59–86.
- PARSLEY, D. C. AND S. WEI (2007): “A prism into the PPP puzzles: The micro-foundations of Big Mac real exchange rates,” *The Economic Journal*, 117, 1336–1356.
- PINGER, P. R. (2017): “Predicting experimental choice behavior and life outcomes from survey measures of present bias,” *Economics Bulletin*, 37, 2162–2172.
- RUSTAGI, D., S. ENGEL, AND M. KOSFELD (2010): “Conditional cooperation and costly monitoring explain success in forest commons management,” *Science*, 330, 961–965.
- SAMUELSON, P. A. (1938): “A note on the pure theory of consumer’s behavior,” *Economica*, 5, 61–71.
- SEN, A. (1973): “Behaviour and the concept of preference,” *Economica*, 40, 241–259.
- SUNDE, U., T. DOHMEN, B. ENKE, A. FALK, D. HUFFMAN, AND G. MEYERHEIM (2022): “Patience and comparative development,” *The Review of Economic Studies*, 89, 2806–2840.



- SUTTER, M., M. G. KOCHER, D. GLÄTZLE-RÜTZLER, AND S. T. TRAUTMANN (2013): “Impatience and uncertainty: Experimental decisions predict adolescents’ field behavior,” *American Economic Review*, 103, 510–531.
- VERSCHOOR, A., B. D’EXELLE, AND B. PEREZ-VIANA (2016): “Lab and life: Does risky choice behaviour observed in experiments reflect that in the real world?” *Journal of Economic Behavior & Organization*, 128, 134–148.
- WEBER, E. U., A. R. BLAIS, AND N. E. BETZ (2002): “A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors,” *Journal of Behavioral Decision Making*, 15, 263–290.



## Appendix

### A Summary Statistics

#### A.1 Study Participants

	China	Iran	Kenya	All
Place of study	Xiamen	Tehran	Nairobi	
Date	Oct 2023	April/May 2023	Nov/Dec 2023	
Female (%)	63.8	53.6	42.6	53.4
Age	19.9 (1.6)	22.7 (3,9)	20.9 (1.9)	21.2 (2.9)
Bachelor Degree	0.92	0.58	0.98	0.83
Master or above	0.08	0.42	0.02	0.17
No. Observations	395	390	383	1168

Table A1: **Demographic Information of Study Participants.** This table shows the time of replication and descriptive statistics for each region as well as for the sum of regions.



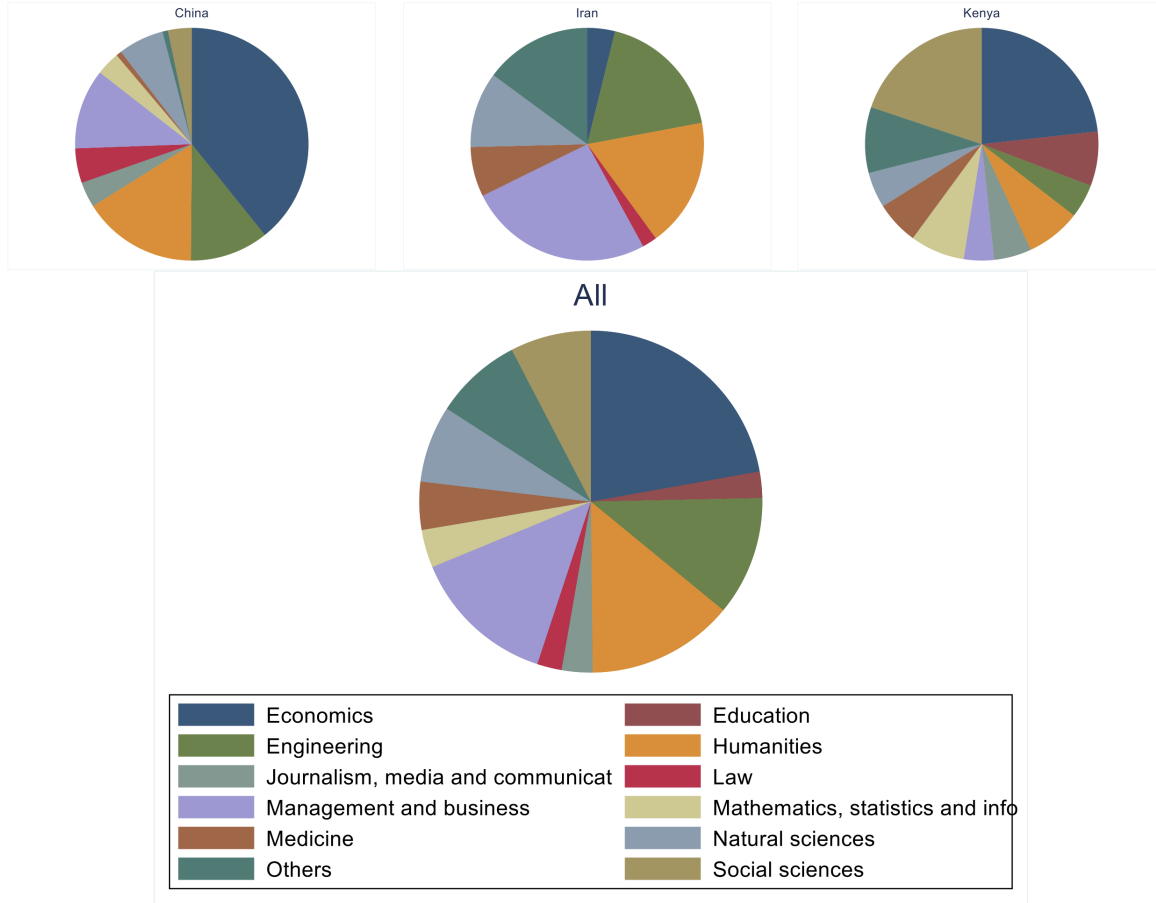


Figure A1: **Distribution of study participants' subject fields.** This pie chart presents the subject fields in each country as well as in the overall sample.

## A.2 Experimental Instruction Comprehension

	China (N=395)	Iran (N=389)	Kenya (N=383)
Risk taking	385 (97%)	341 (88%)	258 (67%)
Time discounting	391 (99%)	370 (95%)	360 (94%)
Trust game	352 (89%)	323 (83%)	317 (83%)
Donation game	378 (96%)	351 (90%)	274 (72%)
Ultimatum game	325 (82%)	238 (61%)	216 (56%)
Prisoners' dilemma	386 (98%)	280 (72%)	242 (63%)

Table A2: **Experimental instruction comprehension.** This table presents the share of study participants who answered the control questions correctly by experiment and country.



### A.3 Summary of Preferences by Country

	China (N=395)			Iran (N=389)			Kenya (N=383)		
	Mean	Median	S.D.	Mean	Median	S.D.	Mean	Median	S.D.
E.M.	9.96	10	2.97	10.54	10.5	3.76	7.64	8.25	5.15
R1	10.21	8	6.09	11.84	12	7.07	7.85	4	8.32
R2	10.89	11	5.42	13.23	13	5.93	8.86	7.67	7.77
R3	4.1	4	1.98	5.17	5	1.86	6.36	7	2.91
R4	3.79	3	1.92	5.19	5	2.11	6.28	7	2.97
R5	4.09	4	1.98	5.03	5	2.21	6.15	7	3.02
R6	4.31	4	2.32	4.52	5	2.14	7.07	8	3.17
R7	3.92	4	2.52	4.4	5	2.44	6.53	8	3.52
R8	4.17	4	2.33	5.04	5	2.3	6.73	8	3.31
R9	5.35	6	2.65	7.18	8	2.28	5.89	6	3.36
R10	2.22	2	2.36	5.32	5	2.55	2.64	1	3.12
R11	4.72	5	2.18	3.89	4	2.39	5.01	5	3.17
R12	5.88	6	2.63	7.92	8	2.13	7.05	8	3
R13	6.12	6	2.34	6.05	6	2.39	7.85	9	2.61
R14	6.81	7	2.16	5.88	6	2.53	5.58	5	3.17
R15	3.43	3	2.01	5.58	6	2.26	5.81	6	3.22
R16	18430.38	20000	16326.84	28187.66	34000	18581.06	28511.75	20000	24763.37

Table A3: **Summary statistics of risk preferences by country.** This table presents the mean, median and standard deviation of survey questions and the experimental measure (E.M.) of risk preferences by country. Items highlighted in blue are GPS survey items.



	China (N=395)			Iran (N=389)			Kenya (N=383)		
	Mean	Median	S.D.	Mean	Median	S.D.	Mean	Median	S.D.
E.M.	11.96	12.5	7.91	6.97	5.75	6.78	2.74	0	5.34
D1	21.44	26	10.45	11.13	8	10.52	6.7	1	9.84
D2	15.49	18	7.78	7.41	6	7.25	5.12	0	7.55
D3	7.19	7	1.82	5.02	5	2.79	7.14	8	3.13
D4	6.9	7	1.87	5.46	6	2.68	7.17	8	3.04
D5	6.44	6	1.72	5.37	6	2.64	7.02	8	2.97
D6	7	7	2.2	5.43	6	2.81	8.25	9	2.66
D7	8.06	8	1.69	6.59	7	2.55	8.19	9	2.56
D8	7.84	8	1.71	6.98	8	2.46	7.98	9	2.87
D9	6.47	7	2.21	5.58	6	2.58	7.17	8	3.02
D10	6.28	7	2.38	5.63	6	2.66	6.51	7	3.25
D11	5.16	6	2.95	4.56	5	3.06	3.72	3	3.61
D12	440.6	20	5859.43	57.35	20	135.48	2639.31	14	51096.22
D13	1071506	3000	14244688.66	5020.93	2500	9938.04	71055.03	2500	1277305.12
D14	8.31	9	1.92	7.2	8	2.69	8.62	10	2.31
D15	6.43	7	2.2	6.06	6	2.52	7.34	8	2.98
D16	4.28	4	2.46	4.39	4	2.57	3.38	2	3.48
D17	3.25	3	2.67	4.2	4	2.99	4.74	5	3.8
D18	4.64	5	2.64	3	2	2.5	2.66	1	3.24
D19	98.82	72	122.52	64.19	25	146.01	107.89	37.5	688.63

Table A4: **Summary statistics of time preferences by country.** This table presents the mean, median and standard deviation of survey questions and the experimental measure of time preferences by country. Items highlighted in blue are GPS survey items.



	China (N=395)			Iran (N=389)			Kenya (N=383)		
	Mean	Median	S.D.	Mean	Median	S.D.	Mean	Median	S.D.
Exp. measure	159.37	125	150.81	127.89	100	138.21	149.8	100	134.4
TR1	6.01	6	1.92	5.38	6	2.17	5.44	5	2.92
TR2	5.96	6	2.15	5.52	6	2.31	5.57	5	2.83
TR3	6.4	7	1.9	5.7	6	2.25	6.11	6	2.71
TR4	5.49	6	1.95	3.68	4	2.28	5.7	6	2.86
TR5	8.05	8	1.37	7.51	8	1.95	7.44	8	2.59
TR6	5.7	6	1.8	5.83	6	2.27	6.33	7	2.66
TR7	3.77	4	1.99	2.42	2	2.1	3.12	2	2.79
TR8	5.62	6	1.95	4.21	4	2.41	5.78	6	2.76
TR9	68.57	40	181.8	47.03	15	101.15	232.54	50	1158.98
TR10	2.77	2	2.8	2.85	2	2.77	3.04	2	2.99
TR11	4.19	4	2.81	2.56	1	2.83	1.25	0	2.29
TR12	2.03	1	2.83	2.18	1	2.7	2.1	0	3.2
TR13	4.55	5	2.32	4.17	4	2.7	3.91	4	3.09
TR14	4.52	5	2.43	3.72	3	2.79	5.64	6	3.16
TR15	6.74	7	2.67	3.63	3	3.06	7.26	9	3.42
TR16	5.33	6	2.26	3.58	3	2.62	3.65	3	3.2
TR17	5.64	6	1.93	4.15	4	2.34	5.43	5	2.96
TR18	4.89	4	2.95	5.47	5	2.58	5.82	6	3.14
TR19	8.54	9	1.41	8.47	9	1.77	8.79	10	2.09
TR20	0.3	0	0.46	0.52	1	0.5	0.71	1	0.45
TR21	0.7	1	0.46	0.48	0	0.5	0.29	0	0.45
TR22	0.31	0	0.46	0.17	0	0.38	0	0	0
TR23	0.69	1	0.46	0.83	1	0.38	0	0	0
TR24	7.52	7	6.16	7	5	6.06	9.05	7.5	7.06

Table A5: **Summary statistics of trust preferences by country.** This table presents the mean, median and standard deviation of survey questions and the experimental measure of trust preferences by country. Items highlighted in blue are GPS survey items.



	China (N=395)			Iran (N=389)			Kenya (N=383)		
	Mean	Median	S.D.	Mean	Median	S.D.	Mean	Median	S.D.
Exp. measure	100.41	100	91.15	89.45	60	88.09	99.16	100	94.36
A1	5.55	6	2.12	6.95	7	2.09	8.25	9	2.46
A2	5.78	6	2.02	7.09	7	2.01	8.05	9	2.5
A3	6.09	6	1.79	7.1	7	1.94	7.98	8	2.22
A4	5.67	6	2	5.66	6	2.44	7.94	9	2.35
A5	7.93	8	1.44	8.22	9	1.8	8.98	10	1.75
A6	5.76	6	1.8	7.24	7	1.99	7.95	9	2.3
A7	4.25	5	2.24	5.36	6	2.58	6.58	7	2.96
A8	5.61	6	1.98	6.06	6	2.46	7.97	9	2.26
A9	7.04	7	2	8.73	9	1.84	8.95	10	1.81
A10	6.66	7	2.35	7.24	8	2.72	8.96	10	1.73
A11	196.01	100	211.2	134.89	100	160.85	258.38	250	224.41
A12	5.73	6	2.33	3.7	3	2.53	3.69	3	3.57
A13	5.88	6	2.34	6.69	7	2.72	8.42	10	2.39
A14	6.27	6	1.95	7.01	7	2.34	8.34	9	2.29
A15	7.65	8	1.69	6.02	6	2.43	6.2	7	3.44
A16	7.39	8	2.07	2.13	2	2.17	2.61	1	3.07
A17	3.03	3	2.45	2.9	2	2.7	3.75	3	3.57
A18	5.39	6	2.01	5.1	5	2.33	7.66	8	2.62
A19	5.41	6	2.09	6.17	6	2.28	6.92	8	3.07
A20	5.42	5	2.24	3.31	3	2.21	2.71	2	3.09
A21	6.37	7	1.73	6.37	7	2.18	7.73	9	2.67
A22	6.67	4	13.29	8.39	3	18.2	99.89	40	153.52
A23	10.17	4	24.08	3.38	1	13.53	55.6	10	170.71

Table A6: **Summary statistics of altruism by country.** This table presents the mean, median and standard deviation of survey questions and the experimental measure of altruism preferences by country. Items highlighted in blue are GPS survey items.



	China (N=395)			Iran (N=389)			Kenya (N=383)		
	Mean	Median	S.D.	Mean	Median	S.D.	Mean	Median	S.D.
Exp. measure	250.47	272.73	166.3	224.19	249.32	161.55	310.62	302.95	191.28
PR1	5.76	6	2.37	8.41	9	1.69	8.12	9	2.46
PR2	6.14	6	2.2	8.25	9	1.66	8.17	9	2.34
PR3	6.15	6	1.88	8.18	8	1.58	8.19	9	2.05
PR4	5.97	6	2.07	6.71	7	2.27	7.74	9	2.57
PR5	7.93	8	1.56	8.95	9	1.48	8.85	10	1.95
PR6	6.15	6	1.95	8.1	8	1.83	8.25	9	2.17
PR7	4.82	5	2.48	6.12	6	2.49	5.87	6	3.11
PR8	6.13	6	2.15	6.9	7	2.28	7.74	8	2.47
PR9	4.5	5	1.14	3.14	4	1.72	3.84	4	1.94
PR10	5.19	5	1.02	4.42	5	1.6	5.15	6	1.88
PR11	8.14	8	1.59	8.6	9	1.75	9	10	1.9
PR12	6.01	6	2.07	8.36	9	1.8	8.72	10	2.2
PR13	7.44	7	1.62	8.13	8	1.82	8.71	10	2.1
PR14	8.52	9	1.61	8.51	9	1.87	7.72	9	3.15
PR15	7.51	8	1.68	7.78	8	1.87	8.31	9	2.37
PR16	17.69	19	8.37	16.75	17.5	7.91	17.24	16.25	11.27

Table A7: **Summary statistics of positive reciprocity by country.** This table presents the mean, median and standard deviation of survey questions and the experimental measure of positive reciprocity preferences by country. Items highlighted in blue are GPS survey items.



	China (N=395)			Iran (N=389)			Kenya (N=383)		
	Mean	Median	S.D.	Mean	Median	S.D.	Mean	Median	S.D.
Exp. measure	162.51	125	113.3	150.07	125	100.14	161.94	130	90.05
NR1	6.33	6	2.26	5.25	5	2.67	3.02	2	2.89
NR2	4.93	5	2.28	5.58	6	2.54	4.57	5	3.2
NR3	5.34	6	2.18	5.44	6	2.5	4.98	5	3.27
NR4	5.15	5	2.03	5.02	5	2.49	4.78	5	3.21
NR5	5.05	5	2.24	4.81	5	2.65	5.28	5	3.36
NR6	4.66	5	2.67	5.14	5	2.8	5.59	6	3.64
NR7	5.4	6	2.25	5.16	5	2.56	5.66	6	3.39
NR8	4.96	5	2.72	4.95	5	2.98	4.62	5	3.42
NR9	4.82	5	2.24	4.85	5	2.54	5.27	5	3.37
NR10	43.6	50	14.33	42.99	50	15.19	45.35	50	19.76
NR11	4.97	5	2.36	4.07	4	2.61	2.59	2	2.92
NR12	5.31	5	2.41	4.23	4	2.68	2.56	1	2.93
NR13	5.9	6	2.56	3.68	3	2.66	2.25	1	2.83
NR14	7.04	7	2.18	3.98	4	2.67	3.01	2	3.26
NR15	7.39	8	2.04	7.49	8	2.33	7.35	8	3.08
NR16	5.53	6	2.75	3.52	3	2.63	2.52	1	2.93
NR17	7.67	8	1.91	6.52	7	2.45	7.18	8	3.26
NR18	6.48	7	2.34	4.76	5	2.78	3.04	2	3.17
NR19	8.13	8	1.96	7.46	8	2.46	8.21	10	2.91
NR20	5.93	6	2.43	8.59	9	1.68	7.67	9	3.07
NR21	4.78	5	2.23	4.22	4	2.72	3.32	2	3.35
NR22	7.67	8	1.95	7.84	8	1.95	8.51	10	2.34

Table A8: **Summary statistics of negative reciprocity by country.** This table presents the mean, median and standard deviation of survey questions and the experimental measure of positive reciprocity preferences by country. Items highlighted in blue are GPS survey items.



In Table A9, we summarize all preferences based on three different sets of data. The left column of Table A9 presents the score of each country on preferences measured by GPS measures among representative samples in China, Iran, Kenya, and Germany, taken from Falk et al. (2018)'s study. The column in the middle shows preferences measured by GPS measures among university students in these four countries. Here, we pooled data from four countries, constructed the GPS measures using the weights applied by Falk et al. (2018), standardized them across countries, and then reported the mean for each country. The right column reports the average of preferences elicited from incentivized lab experiments with the same university students in these four countries, with the country means calculated in the same way as those in the middle.

	<b>Falk et al. (2018) GPS</b>	<b>This study GPS</b>	<b>This study Experiment</b>
<b>Risk</b>	DE (-0.044)	CN (-0.114)	KE (-0.497)
	CN (-0.020)	KE (-0.095)	CN (0.084)
	KE (0.244)	DE (-0.017)	DE (0.170)
	IR (0.338)	IR (0.227)	IR (0.230)
<b>Time</b>	IR (-0.381)	KE (-0.521)	KE (-0.769)
	KE (-0.076)	IR (-0.225)	IR (-0.263)
	CN (0.398)	DE (0.026)	CN (0.333)
	DE (0.624)	CN (0.700)	DE (0.669)
<b>Trust</b>	KE (-0.422)	IR (-0.267)	IR (-0.182)
	DE (-0.119)	KE (-0.244)	KE (-0.030)
	IR (0.300)	DE (0.157)	CN (0.036)
	CN (0.553)	CN (0.342)	DE (0.172)
<b>Altruism</b>	KE (-0.317)	DE (-0.314)	IR (-0.176)
	DE (-0.051)	IR (-0.207)	KE (-0.076)
	CN (0.505)	CN (0.098)	CN (-0.063)
	IR (0.590)	KE (0.435)	DE (0.310)
<b>Pos. Reciprocity</b>	KE (-0.318)	DE (-0.260)	IR (-0.196)
	DE (-0.080)	IR (-0.168)	DE (-0.055)
	CN (0.479)	CN (0.179)	CN (-0.045)
	IR (0.551)	KE (0.239)	KE (0.302)
<b>Neg. Reciprocity</b>	DE (-0.294)	–	DE (-1.059)
	KE (-0.096)	KE (-0.522)	IR (0.285)
	CN (0.026)	IR (0.153)	KE (0.391)
	IR (0.119)	CN (0.355)	CN (0.396)

Table A9: **Measures by preference and country.** This table presents rankings and averages (in parentheses) based on the GPS measure from Falk et al. (2018), the GPS measure from this study and the experimental measure from this study for each country and preference.



## B Replication Tables

In these tables, we use red to indicate insignificant coefficients, and we highlight the results in bold wherever the 95% confidence interval test is passed. Each table is presented once with the full sample and a second time with the smaller sample of participants who answered the control questions correctly.



Preference	Items	Germany	China	Iran	Kenya	Rep. combined, FE	Rep. combined, no FE
Risk	Staircase measure (R1)	0.216*** (0.054)	<b>0.266***</b> <b>(0.060)</b> [0.148,0.385]	<b>0.230***</b> <b>(0.051)</b> [0.129,0.331]	<b>0.298***</b> <b>(0.052)</b> [0.196,0.400]	<b>0.269***</b> <b>(0.032)</b> [0.206,0.332]	<b>0.342***</b> <b>(0.031)</b> [0.280,0.403]
	Willingness to take risks (R3)	0.241*** (0.051)	0.050 <b>(0.060)</b> [-0.069,0.169]	0.125*** (0.046) [0.034,0.216]	0.029 <b>(0.054)</b> [-0.078,0.136]	0.055 <b>(0.036)</b> [-0.015,0.124]	-0.040 <b>(0.032)</b> [-0.103,0.023]
	Adjusted R-squared	0.140	0.080	0.084	0.088	0.163	0.111
Time	Staircase measure (D1)	0.442*** (0.049)	<b>0.417***</b> <b>(0.051)</b> [0.317,0.517]	0.219*** (0.054) [0.112,0.325]	0.190*** (0.065) [0.062,0.318]	0.312*** (0.032) [0.249,0.375]	<b>0.497***</b> <b>(0.027)</b> [0.443,0.551]
	Willingness to wait (D3)	0.179*** (0.050)	<b>0.130**</b> <b>(0.052)</b> [0.029,0.232]	<b>0.180***</b> <b>(0.052)</b> [0.078,0.282]	0.056 <b>(0.047)</b> [-0.037,0.149]	0.101*** (0.025) [0.051,0.151]	0.025 <b>(0.024)</b> [-0.023,0.072]
	Adjusted R-squared	0.299	0.238	0.111	0.039	0.333	0.254
Trust	People have only the best intentions. (T16)	0.266*** (0.048)	0.166*** (0.050) [0.068,0.263]	<b>0.182***</b> <b>(0.050)</b> [0.084,0.280]	<b>0.191***</b> <b>(0.053)</b> [0.088,0.295]	0.184*** (0.030) [0.125,0.243]	0.189*** (0.029) [0.133,0.246]
	Adjusted R-squared	0.070	0.025	0.031	0.034	0.037	0.035
Altruism	Hypothetical donation game (A11)	0.185*** (0.049)	<b>0.243***</b> <b>(0.052)</b> [0.141,0.346]	<b>0.262***</b> <b>(0.060)</b> [0.144,0.380]	<b>0.175***</b> <b>(0.059)</b> [0.059,0.290]	<b>0.228***</b> <b>(0.034)</b> [0.161,0.295]	<b>0.225***</b> <b>(0.034)</b> [0.159,0.292]
	Share when it comes to good causes (A10)	0.321*** (0.044)	<b>0.281***</b> <b>(0.045)</b> [0.192,0.370]	0.103*** (0.054) [-0.002,0.209]	-0.053 <b>(0.064)</b> [-0.179,0.073]	0.142*** (0.034) [0.076,0.208]	0.109*** (0.030) [0.051,0.167]
	Adjusted R-squared	0.175	0.175	0.094	0.025	0.081	0.076
Pos. Reciprocity	When someone does me a favor (PR11)	0.271*** (0.050)	0.103** (0.052) [0.001,0.205]	0.164*** (0.045) [0.075,0.252]	0.023 <b>(0.057)</b> [-0.089,0.136]	0.092*** (0.032) [0.028,0.156]	0.117*** (0.033) [0.051,0.182]
	Thank-you gift in the lost in the city scenario (PR9)	0.288*** (0.050)	<b>0.201***</b> <b>(0.047)</b> [0.108,0.295]	0.163*** (0.051) [0.062,0.264]	0.036 <b>(0.052)</b> [-0.067,0.139]	0.115*** (0.031) [0.053,0.176]	0.120*** (0.030) [0.062,0.179]
	Adjusted R-squared	0.194	0.058	0.062	-0.003	0.064	0.033
Neg. Reciprocity	Punish someone who treats you unfairly (NR1)	N.A.	0.190*** (0.069) [0.054,0.325]	-0.044 <b>(0.088)</b> [-0.217,0.129]	0.096 <b>(0.069)</b> [-0.040,0.232]	0.106** (0.046) [0.016,0.196]	0.073* (0.042) [-0.008,0.155]
	Punish someone who treats others unfairly (NR2)	N.A.	-0.089 <b>(0.063)</b> [-0.212,0.035]	0.074 <b>(0.063)</b> [-0.050,0.198]	-0.029 <b>(0.064)</b> [-0.155,0.098]	-0.015 <b>(0.036)</b> [-0.086,0.056]	-0.012 <b>(0.035)</b> [-0.082,0.057]
	Take revenge at the first occasion (NR11)	N.A.	0.076 <b>(0.067)</b> [-0.055,0.207]	0.091 <b>(0.085)</b> [-0.075,0.258]	-0.018 <b>(0.054)</b> [-0.125,0.089]	0.038 <b>(0.039)</b> [-0.038,0.114]	0.036 <b>(0.038)</b> [-0.039,0.111]
	Adjusted R-squared	N.A.	0.038	0.004	-0.001	0.012	0.007
	Number of observations	382 (360)	395	389	383	1167	1167

Table B1: **Regression coefficients of GPS items.** Coefficients from OLS regressions with robust standard errors in parentheses and 95% confidence intervals in square brackets. Blue colored items are GPS items. Germany is the reference column. Black with boldface highlights coefficients that are statistically indistinguishable from that in Germany. Plain black indicates coefficients that are statistically different from that in Germany. Red indicates insignificant coefficients. For negative reciprocity, we cannot compare because the original German sample does not have GPS items NR1 and NR2. \*\*\* p<0.01; \*\*p<0.05; \*p<0.1.



Preference	Items	Germany	China	Iran	Kenya	Rep. combined FE	Rep. combined, no FE
Risk	Staircase measure (R1)	0.216*** (0.054)	<b>0.256***</b> (0.061)	<b>0.234***</b> (0.057)	<b>0.324***</b> (0.065)	<b>0.277***</b> (0.038)	<b>0.350***</b> (0.037)
	Willingness to take risks (R3)	0.241*** (0.051)	0.064 (0.061)	0.112** (0.051)	<b>0.112*</b> (0.066)	0.097** (0.040)	-0.017 (0.037)
			[-0.055,0.184]	[0.010,0.213]	[-0.019,0.242]	[0.019,0.175]	[-0.091,0.056]
	Adjusted R-squared	0.140	0.079	0.079	0.113	0.164	0.108
Time	Staircase measure (D1)	0.442*** (0.049)	<b>0.413***</b> (0.051)	0.192*** (0.055)	0.186*** (0.068)	0.273*** (0.029)	<b>0.449***</b> (0.025)
	Willingness to wait (D3)	0.179*** (0.050)	<b>0.123**</b> (0.052)	<b>0.212***</b> (0.051)	0.056 (0.048)	0.108*** (0.026)	0.030 (0.025)
			[-0.021,0.225]	[-0.111,0.313]	[-0.038,0.150]	[0.057,0.158]	[-0.019,0.078]
	Adjusted R-squared	0.299	0.229	0.115	0.039	0.345	0.259
Trust	People have only the best intentions. (T16)	0.266*** (0.048)	<b>0.189***</b> (0.053)	<b>0.223***</b> (0.057)	0.140** (0.056)	0.186*** (0.033)	0.194*** (0.031)
			[-0.084,0.293]	[-0.111,0.334]	[0.029,0.250]	[0.121,0.250]	[0.133,0.255]
	Adjusted R-squared	0.070	0.032	0.042	0.018	0.035	0.036
Altruism	Hypothetical donation game (A11)	0.185*** (0.049)	<b>0.245***</b> (0.053)	<b>0.261***</b> (0.063)	<b>0.230***</b> (0.064)	<b>0.224***</b> (0.032)	<b>0.221***</b> (0.031)
	Share when it comes to good causes (A10)	0.321*** (0.044)	<b>0.285***</b> (0.046)	0.102* (0.058)	-0.037 (0.071)	0.145*** (0.032)	0.115*** (0.028)
			[-0.194,0.375]	[-0.012,0.217]	[-0.176,0.102]	[0.082,0.208]	[0.060,0.171]
	Adjusted R-squared	0.175	0.183	0.091	0.055	0.107	0.101
Pos. Reciprocity	When someone does me a favor (PR11)	0.271*** (0.050)	0.096* (0.054)	<b>0.196***</b> (0.054)	0.014 (0.060)	0.099*** (0.037)	0.122*** (0.037)
	Thank-you gift in the lost in the city scenario (PR9)	0.288*** (0.050)	<b>0.210***</b> (0.052)	<b>0.179***</b> (0.058)	0.039 (0.055)	0.125*** (0.034)	0.121*** (0.033)
			[-0.106,0.313]	[-0.065,0.292]	[-0.069,0.148]	[0.058,0.192]	[0.056,0.186]
	Adjusted R-squared	0.194	0.058	0.073	-0.004	0.051	0.035
Neg. Reciprocity	Punish someone who treats you unfairly (NR1)	N.A.	0.160** (0.077)	-0.188 (0.145)	0.099 (0.084)	0.082 (0.059)	0.089 (0.055)
			[-0.009,0.311]	[-0.475,0.099]	[-0.067,0.266]	[-0.034,0.197]	[-0.018,0.196]
	Punish someone who treats others unfairly (NR2)	N.A.	-0.081 (0.070)	0.161 (0.106)	-0.070 (0.072)	-0.014 (0.048)	-0.024 (0.047)
			[-0.219,0.057]	[-0.047,0.370]	[-0.211,0.072]	[-0.109,0.081]	[-0.117,0.069]
	Take revenge at the first occasion (NR11)	N.A.	0.059 (0.074)	0.228* (0.124)	0.043 (0.081)	0.077 (0.052)	0.078 (0.051)
			[-0.086,0.203]	[-0.017,0.474]	[-0.117,0.203]	[-0.024,0.179]	[-0.023,0.178]
	Adjusted R-squared	N.A.	0.020	0.029	-0.004	0.019	0.018
Number of observations		382 (360)	321-391	186-370	151-360	658-1121	658-1121

Table B2: **GPS Items - Restricted solely to participants who passed control questions.** Coefficients from OLS regressions with robust standard errors in parentheses and 95% confidence intervals in square brackets. Blue colored items are GPS items. Germany is the reference column. Black with boldface highlights coefficients that are statistically indistinguishable from that in Germany. Plain black indicates coefficients that are statistically different from that in Germany. Red indicates insignificant coefficients. For negative reciprocity, we cannot compare because the original German sample does not have GPS items NR1 and NR2. \*\*\* p<0.01; \*\*p<0.05; \*p<0.1.



Preference	Items	Germany	China	Iran	Kenya	Rep. combined, FE	Rep. combined, no FE
Risk	Staircase measure (R1)	0.353***	<b>0.288***</b>	<b>0.294***</b>	<b>0.296***</b>	0.282***	<b>0.303***</b>
	Willingness to take risks (R3)	(0.052)	<b>0.054</b> [0.183,0.394]	<b>0.049</b> [0.197,0.391]	<b>0.050</b> [0.197,0.394]	(0.030) [0.224,0.340]	<b>(0.020)</b> [0.245,0.362]
	Adjusted R-squared	0.124	0.081	0.084	0.085	0.162	0.091
Time	Staircase measure (D1)	0.539***	<b>0.484***</b>	0.313***	0.208***	0.357***	<b>0.505***</b>
	Willingness to wait (D3)	(0.044)	<b>(0.042)</b> [0.401,0.567]	(0.052) [0.210,0.416]	(0.066) [0.078,0.338]	(0.031) [0.297,0.418]	<b>0.029</b> [0.453,0.558]
	Adjusted R-squared	0.289	0.232	0.096	0.041	0.330	0.255
Trust	People have only the best intentions. (T16)	0.266***	0.166***	<b>0.182***</b>	<b>0.191***</b>	0.184***	0.189***
		(0.048)	(0.050) [0.068,0.263]	<b>(0.050)</b> [0.084,0.280]	<b>(0.053)</b> [0.088,0.295]	(0.030) [0.125,0.242]	(0.028) [0.133,0.245]
	Adjusted R-squared	0.070	0.025	0.031	0.034	0.037	0.035
Altruism	Hypothetical donation game (A11)	0.321***	<b>0.328***</b>	<b>0.298***</b>	0.165***	<b>0.267***</b>	<b>0.261***</b>
	Share when it comes to good causes (A10)	(0.047)	<b>(0.047)</b> [0.231,0.426]	(0.044) [0.188,0.409]	(0.052) [0.051,0.280]	<b>0.031</b> [0.201,0.332]	<b>0.032</b> [0.197,0.324]
	Adjusted R-squared	0.091	0.106	0.087	0.025	0.067	0.067
Pos. Reciprocity	When someone does me a favor (PR11)	0.438***	0.237***	0.259***	0.048	0.163***	0.187***
	Thank-you gift in the lost in the city scenario (PR9)	(0.047)	(0.047) [0.145,0.329]	(0.044) [0.171,0.346]	<b>(0.052)</b> [-0.055,0.151]	(0.031) [0.102,0.223]	(0.032) [0.124,0.249]
	Adjusted R-squared	0.195	0.054	0.065	-0.000	0.065	0.034
Neg. Reciprocity	Punish someone who treats you unfairly (NR1)	NA	0.160***	0.095*	0.042	0.107***	0.086***
	Punish someone who treats others unfairly (NR2)		(0.049)	(0.050)	<b>(0.053)</b>	(0.031)	(0.029)
	Take revenge at the first occasion (NR11)		[0.064,0.257]	[-0.003,0.194]	[-0.062,0.146]	[0.046,0.167]	[0.029,0.143]
	Adjusted R-squared		0.023	0.007	-0.001	0.010	0.007
Number of observations		382 (360)	395	389	383	1167	1167

Table B3: **Regression coefficients of GPS measures.** Coefficients from OLS regressions with robust standard errors in parentheses and 95% confidence intervals in square brackets. Blue colored items are GPS items. Germany is the reference column. Black with boldface highlights coefficients that are statistically indistinguishable from that in Germany. Plain black indicates coefficients that are statistically different from that in Germany. Red indicates insignificant coefficients. For negative reciprocity, we cannot compare because the original German sample does not have GPS items NR1 and NR2. \*\*\* p<0.01; \*\*p<0.05; \*p<0.1.



Preference	Items	Germany	China	Iran	Kenya	Rep. combined FE	Rep. combined, no FE
Risk	Staircase measure (R1)	0.353***	<b>0.287***</b>	<b>0.292***</b>	<b>0.363***</b>	<b>0.300***</b>	<b>0.310***</b>
	Willingness to take risks (R3)	(0.052)	( <b>0.054</b> )	( <b>0.054</b> )	( <b>0.062</b> )	( <b>0.032</b> )	( <b>0.033</b> )
			[0.181,0.394]	[0.186,0.397]	[0.241,0.486]	[0.236,0.363]	[0.246,0.374]
	Adjusted R-squared	0.124	0.081	0.081	0.117	0.165	0.095
Time	Staircase measure (D1)	0.539***	<b>0.475***</b>	0.303***	0.204***	0.353***	<b>0.510***</b>
	Willingness to wait (D3)	(0.040)	( <b>0.043</b> )	(0.053)	(0.070)	(0.032)	( <b>0.027</b> )
			[0.392,0.559]	[0.197,0.408]	[0.068,0.341]	[0.291,0.415]	[0.456,0.563]
	Adjusted R-squared	0.289	0.225	0.092	0.040	0.341	0.259
Trust	People have only the best intentions. (T16)	0.266***	<b>0.189***</b>	<b>0.223***</b>	0.140**	0.187***	0.195***
		(0.048)	( <b>0.053</b> )	( <b>0.057</b> )	(0.056)	(0.033)	(0.031)
			[0.084,0.293]	[0.111,0.334]	[0.029,0.250]	[0.122,0.251]	[0.134,0.256]
	Adjusted R-squared	0.070	0.032	0.042	0.018	0.035	0.036
Altruism	Hypothetical donation game (A11)	0.321***	<b>0.333***</b>	<b>0.295***</b>	<b>0.223***</b>	<b>0.296***</b>	<b>0.289***</b>
	Share when it comes to good causes (A10)	(0.048)	( <b>0.050</b> )	( <b>0.059</b> )	( <b>0.064</b> )	( <b>0.034</b> )	( <b>0.033</b> )
			[0.235,0.432]	[0.178,0.411]	[0.097,0.349]	[0.228,0.363]	[0.224,0.355]
	Adjusted R-squared	0.091	0.111	0.084	0.057	0.089	0.089
Pos. Reciprocity	When someone does me a favor (PR11)	0.438***	0.235***	0.296***	0.044	0.175***	0.191***
	Thank-you gift in the lost in the city scenario (PR9)	(0.047)	(0.050)	(0.053)	( <b>0.057</b> )	(0.035)	(0.036)
			[0.137,0.334]	[0.191,0.401]	[-0.068,0.157]	[0.107,0.244]	[0.121,0.261]
	Adjusted R-squared	0.195	0.053	0.076	-0.001	0.051	0.036
Neg. Reciprocity	Punish someone who treats you unfairly (NR1)	NA	0.124**	0.157**	0.064	0.133***	0.140***
	Punish someone who treats others unfairly (NR2)		(0.055)	(0.069)	( <b>0.075</b> )	(0.042)	(0.038)
	Take revenge at the first occasion (NR11)		[0.015,0.233]	[0.020,0.293]	[-0.085,0.212]	[0.051,0.214]	[0.065,0.214]
	Adjusted R-squared		0.012	0.019	-0.000	0.019	0.017
Number of observations		382 (360)	321-391	186-370	151-360	658-1121	658-1121

Table B4: **GPS Measures - Restricted solely to participants who passed control questions.** Coefficients from OLS regressions with robust standard errors in parentheses and 95% confidence intervals in square brackets. Blue colored items are GPS items. Germany is the reference column. Black with boldface highlights coefficients that are statistically indistinguishable from that in Germany. Plain black indicates coefficients that are statistically different from that in Germany. Red indicates insignificant coefficients. For negative reciprocity, we cannot compare because the original German sample does not have GPS items NR1 and NR2. \*\*\* p<0.01; \*\*p<0.05; \*p<0.1.



Preference	Items	Germany	China	Iran	Kenya	Rep. combined, FE	Rep. combined, no FE
Risk	Multiple price lists (R2)	0.276*** (0.057)	<b>0.335***</b> <b>(0.057)</b> [0.223,0.446]	<b>0.261***</b> <b>(0.055)</b> [0.153,0.370]	<b>0.413***</b> <b>(0.048)</b> [0.319,0.507]	<b>0.357***</b> <b>(0.031)</b> [0.295,0.418]	<b>0.425***</b> <b>(0.030)</b> [0.366,0.484]
	Willingness to take risk (R3)	0.203*** (0.049)	0.016 <b>(0.061)</b> [-0.104,0.136]	0.092* (0.049)	0.007 <b>(0.051)</b> [-0.094,0.107]	0.021 <b>(0.034)</b> [-0.046,0.087]	-0.066** <b>(0.028)</b> [-0.126,-0.006]
	Adjusted R-squared	0.162	0.112	0.093	0.167	0.206	0.171
Time	Multiple price lists (D2)	0.485*** (0.052)	<b>0.470***</b> <b>(0.049)</b> [0.374,0.565]	0.192*** (0.057)	0.186*** (0.062)	0.371*** (0.032)	<b>0.499***</b> <b>(0.028)</b> [0.443,0.554]
	Willingness to wait in comparison with others? (D4)	0.171*** (0.050)	0.076 <b>(0.049)</b> [-0.021,0.173]	<b>0.195***</b> <b>(0.054)</b> [0.089,0.302]	-0.006 <b>(0.044)</b> [-0.094,0.081]	0.075*** (0.025)	-0.008 <b>(0.025)</b> [-0.057,0.041]
	Adjusted R-squared	0.340	0.256	0.101	0.029	0.328	0.245
Trust	First-mover behavior in trust game (T24)	0.629*** (0.043)	<b>0.570***</b> <b>(0.051)</b> [0.470,0.669]	0.471*** (0.057)	0.310*** (0.054)	0.447*** 0.032	0.444*** 0.031
	People have only the best intentions. (T16)	0.133*** (0.038)	0.057 <b>(0.040)</b> [-0.021,0.135]	<b>0.108**</b> <b>(0.046)</b> [0.017,0.199]	<b>0.162***</b> <b>(0.049)</b> [0.066,0.258]	<b>0.120***</b> <b>(0.027)</b> [0.066,0.172]	<b>0.135***</b> <b>(0.025)</b> [0.084,0.184]
	Adjusted R-squared	0.452	0.337	0.245	0.127	0.230	0.229
Altruism	Hypothetical donation game (A11)	0.185*** (0.049)	<b>0.243***</b> <b>(0.052)</b> [0.141,0.346]	<b>0.262***</b> <b>(0.060)</b> [0.144,0.380]	<b>0.175***</b> <b>(0.059)</b> [0.059,0.290]	<b>0.228***</b> <b>(0.034)</b> [0.161,0.295]	<b>0.225***</b> <b>(0.034)</b> [0.159,0.292]
	Share when it comes to good causes. (A10)	0.321*** (0.044)	<b>0.281***</b> <b>(0.045)</b> [0.192,0.370]	0.103* (0.054)	-0.053 <b>(0.064)</b> [-0.179,0.073]	0.142*** (0.034)	0.109*** (0.030)
	Adjusted R-squared	0.175	0.175	0.094	0.025	0.081	0.076
Pos. Reciprocity	Second-mover behavior in trust game (PR16)	0.486*** (0.049)	<b>0.506***</b> <b>(0.045)</b> [0.418,0.593]	<b>0.399***</b> <b>(0.049)</b> [0.302,0.496]	0.257*** (0.060)	0.359*** (0.034)	0.357*** (0.033)
	Thank-you gift in the lost in the city scenario (PR9)	0.164*** (0.049)	0.069* (0.042)	<b>0.098*</b> <b>(0.050)</b> [-0.013,0.151]	-0.028 <b>(0.052)</b> [-0.130,0.074]	0.037 <b>(0.031)</b> [-0.023,0.097]	0.049* (0.030)
	Adjusted R-squared	0.329	0.279	0.185	0.058	0.176	0.138
Neg. Reciprocity	Minimum acceptable offer (NR10)	0.340*** (0.058)	0.226*** (0.046)	0.121** (0.055)	0.054 <b>(0.061)</b> [-0.065,0.174]	0.129*** (0.031)	0.131*** (0.031)
	Adjusted R-squared	0.115	0.049	0.012	0.000	0.017	0.016
Number of observations		382 (360)	395	389	383	1167	1167

Table B5: **Regression coefficients of PSM items.** Coefficients from OLS regressions with robust standard errors in parentheses and 95% confidence internals in square brackets. Blue colored items are GPS items. Germany is the reference column. Black with boldface highlights coefficients that are statistically indistinguishable from that in Germany. Plain black indicates coefficients that are statistically different from that in Germany. Red indicates insignificant coefficients or significant coefficients in the opposite direction. For negative reciprocity we only include one quantitative item because the original one qualitative question on willingness to punish people has been split into two questions for the GPS. \*\*\* p<0.01; \*\*p<0.05; \*p<0.1.



Preference	Items	Germany	China	Iran	Kenya	Rep. combined FE	Rep. combined no FE
Risk	Multiple price lists (R2)	0.276*** (0.057)	<b>0.330***</b> (0.058)	<b>0.266***</b> (0.059)	<b>0.456***</b> (0.061)	<b>0.365***</b> (0.036)	<b>0.436***</b> (0.035)
	Willingness to take risk (R3)	0.203*** (0.049)	<b>0.216,0.444</b> 0.027 (0.062)	<b>0.150,0.383</b> 0.073 (0.055)	<b>0.337,0.576</b> 0.061 (0.064)	<b>0.295,0.435</b> 0.046 (0.039)	<b>0.368,0.504</b> -0.063* (0.036)
			<b>[-0.094,0.148]</b>	<b>[-0.035,0.181]</b>	<b>[-0.066,0.187]</b>	<b>[-0.031,0.122]</b>	<b>[-0.134,0.008]</b>
	Adjusted R-squared	0.162	0.112	0.090	0.194	0.204	0.164
Time	Multiple price lists (D2)	0.485*** (0.052)	<b>0.471***</b> (0.049)	0.175*** (0.058)	0.185*** (0.062)	0.289*** (0.030)	<b>0.462***</b> (0.026)
	Willingness to wait in comparison with others? (D4)	0.171*** (0.050)	<b>0.375,0.566</b> 0.061 (0.050)	<b>0.234***</b> (0.055)	<b>-0.024</b> (0.045)	0.068*** (0.023)	<b>-0.009</b> (0.023)
			<b>[-0.037,0.160]</b>	<b>[0.126,0.342]</b>	<b>[-0.113,0.065]</b>	<b>[0.022,0.113]</b>	<b>[-0.054,0.037]</b>
	Adjusted R-squared	0.340	0.250	0.112	0.029	0.341	0.250
Trust	First-mover behavior in trust game (T24)	0.629*** (0.043)	<b>0.557***</b> (0.054)	0.487*** (0.063)	0.335*** (0.058)	1.267*** (0.094)	1.244*** (0.093)
	People have only the best intentions. (T16)	0.133*** (0.038)	<b>0.451,0.663</b> <b>0.081*</b> (0.043)	<b>0.364,0.610</b> <b>0.128**</b> (0.052)	<b>0.221,0.450</b> <b>0.102**</b> (0.052)	<b>1.083,1.451</b> <b>0.109***</b> (0.029)	<b>1.061,1.428</b> <b>0.129***</b> (0.027)
			<b>[-0.003,0.166]</b>	<b>[0.026,0.230]</b>	<b>[0.000,0.203]</b>	<b>[0.052,0.167]</b>	<b>[0.076,0.183]</b>
	Adjusted R-squared	0.452	0.329	0.267	0.136	0.244	0.240
Altruism	Hypothetical donation game (A11)	0.185*** (0.049)	<b>0.245***</b> (0.053)	<b>0.261***</b> (0.063)	<b>0.230***</b> (0.064)	<b>0.224***</b> (0.032)	<b>0.221***</b> (0.031)
	Share when it comes to good causes. (A10)	0.321*** (0.044)	<b>0.140,0.350</b> <b>0.285***</b> (0.046)	<b>0.138,0.385</b> 0.102* (0.058)	<b>0.103,0.357</b> <b>-0.037</b> (0.071)	<b>0.161,0.286</b> 0.145*** (0.032)	<b>0.159,0.282</b> 0.115*** (0.028)
			<b>[0.194,0.375]</b>	<b>[-0.012,0.217]</b>	<b>[-0.176,0.102]</b>	<b>[0.082,0.208]</b>	<b>[0.060,0.171]</b>
	Adjusted R-squared	0.175	0.183	0.091	0.055	0.107	0.101
Pos. Reciprocity	Second-mover behavior in trust game (PR16)	0.486*** (0.049)	<b>0.494***</b> (0.048)	<b>0.392***</b> (0.056)	0.242*** (0.067)	0.363*** (0.037)	0.359*** (0.036)
	Thank-you gift in the lost in the city scenario (PR9)	0.164*** (0.049)	<b>0.399,0.589</b> 0.073 (0.047)	<b>0.282,0.502</b> <b>0.114**</b> (0.057)	<b>0.110,0.374</b> <b>-0.022</b> (0.055)	<b>0.290,0.436</b> <b>0.045</b> (0.034)	<b>0.289,0.430</b> <b>0.048</b> (0.033)
			<b>[-0.019,0.166]</b>	<b>[0.002,0.227]</b>	<b>[-0.131,0.087]</b>	<b>[-0.021,0.111]</b>	<b>[-0.015,0.112]</b>
	Adjusted R-squared	0.329	0.264	0.180	0.053	0.162	0.138
Neg. Reciprocity	Minimum acceptable offer (NR10)	0.340*** (0.058)	<b>0.245***</b> (0.049)	0.122* (0.067)	<b>0.010</b> (0.118)	0.164*** (0.039)	0.160*** (0.038)
			<b>[0.148,0.342]</b>	<b>[-0.010,0.254]</b>	<b>[-0.223,0.244]</b>	<b>[0.086,0.241]</b>	<b>[0.085,0.235]</b>
	Adjusted R-squared	0.115	0.057	0.011	-0.007	0.032	0.025
Number of observations		382 (360)	321-391	186-370	151-360	658-1121	658-1121

Table B6: **PSM - Restricted solely to participants who passed control questions.** Coefficients from OLS regressions with robust standard errors in parentheses and 95% confidence intervals in square brackets. Blue colored items are GPS items. Germany is the reference column. Black with boldface highlights coefficients that are statistically indistinguishable from that in Germany. Plain black indicates coefficients that are statistically different from that in Germany. Red indicates insignificant coefficients or significant coefficients in the opposite direction. For negative reciprocity we only include one quantitative item because the original one qualitative question on willingness to punish people has been split into two questions for the GPS. \*\*\* p<0.01; \*\*p<0.05; \*p<0.1.



## **C Additional Results and Details**

### **C.1 Consistency Bias**

See the table on the next page.



Preference	Items	China		Iran		Kenya		All Rep.	All Rep.
		Corr	Coef.	Corr	Coef.	Corr	Coef.	Corr	Coef.
Risk	Exp. first	0.418***	0.415*** (0.069)	0.193***	0.268*** (0.080)	0.481***	0.473*** (0.063)	0.426***	0.409*** (0.041)
	Survey first	0.307***	0.250*** (0.071)	0.351***	0.327*** (0.065)	0.362***	0.346*** (0.071)	0.361***	0.314*** (0.041)
	P-value		0.106		0.569		0.183		0.098
Time	Exp. first	0.607***	0.574*** (0.059)	0.340***	0.335*** (0.074)	0.504***	0.441*** (0.101)	0.622***	0.423*** (0.039)
	Survey first	0.489***	0.447*** (0.060)	0.226***	0.229*** (0.076)	0.173**	0.082 (0.077)	0.389***	0.291*** (0.040)
	P-value		0.133		0.318		0.006		0.008
Trust	Exp. first	0.699***	0.664*** (0.061)	0.500***	0.418*** (0.079)	0.331***	0.381*** (0.078)	0.517***	0.494*** (0.043)
	Survey first	0.452***	0.495*** (0.077)	0.472***	0.566*** (0.082)	0.217***	0.266*** (0.086)	0.394***	0.432*** (0.048)
	P-value		0.089		0.206		0.353		0.350
Altruism	Exp. first	0.397***	0.388*** (0.067)	0.387***	0.388*** (0.067)	0.165**	0.181*** (0.077)	0.312***	0.308*** (0.044)
	Survey first	0.474***	0.276*** (0.068)	0.239***	0.200*** (0.087)	0.212***	0.145 (0.088)	0.309***	0.217*** (0.047)
	P-value		0.242		0.087		0.754		0.153
Pos. Reciprocity	Exp. first	0.664***	0.618*** (0.045)	0.422***	0.407*** (0.061)	0.389***	0.389*** (0.078)	0.457***	0.443*** (0.041)
	Survey first	0.419***	0.438*** (0.069)	0.426***	0.449*** (0.070)	0.145*	0.064 (0.063)	0.330***	0.278*** (0.046)
	P-value		0.031		0.649		0.002		0.008
Neg. Reciprocity	Exp. first	0.112	0.153** (0.068)	0.144**	0.040 (0.078)	0.081	0.121 (0.079)	0.116***	0.094** (0.042)
	Survey first	0.254***	0.297*** (0.062)	0.238***	0.231*** (0.059)	0.012	-0.061 (0.093)	0.168***	0.179*** (0.046)
	P-value		0.118		0.051		0.140		0.176
Number of observations			395		389		383		1167

Table C1: **Consistency bias.** We report spearman correlation between behavior in incentivized games and the corresponding hypothetical version by order (experiment first or survey first). We further report univariate coefficients from OLS regressions for each country and all replication countries combined, with robust standard errors in parentheses. In the all replication countries combined sample, we controlled for country fixed effects. Red-colored p-value shows significant order effects. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1.



## C.2 Correlations

		Risk	Time	Trust	Altruism	Pos. reciprocity	Neg. reciprocity
Panel A: Experimental measure	Risk	1.00					
	Time	<b>0.19***</b>	1.00				
	Trust	0.07***	0.04	1.00			
	Altruism	0.07***	0.03	<b>0.26***</b>	1.00		
	Pos. reciprocity	-0.09***	-0.07***	<b>0.47***</b>	<b>0.27***</b>	1.00	
	Neg. reciprocity	-0.06**	<b>-0.18***</b>	-0.08***	-0.09***	0.01	1.00
Panel B: GPS qualitative question	Risk	1.00					
	Time	<b>0.18***</b>	1.00				
	Trust	-0.05**	0.01	1.00			
	Altruism	<b>0.19***</b>	0.10***	0.09***	1.00		
	Pos. reciprocity	0.09***	0.05*	<b>0.12***</b>	<b>0.41***</b>	1.00	
	Neg. reciprocity	-0.06**	-0.04	0.03	-0.10***	-0.02	1.00

Table C2: **Correlations.** This table presents the correlations between the GPS measures and GPS qualitative questions with each preference. \*\*\* p<0.01; \*\*p<0.05; \*p<0.1.



### C.3 Adjusted $R^2$ Plots in Different Countries

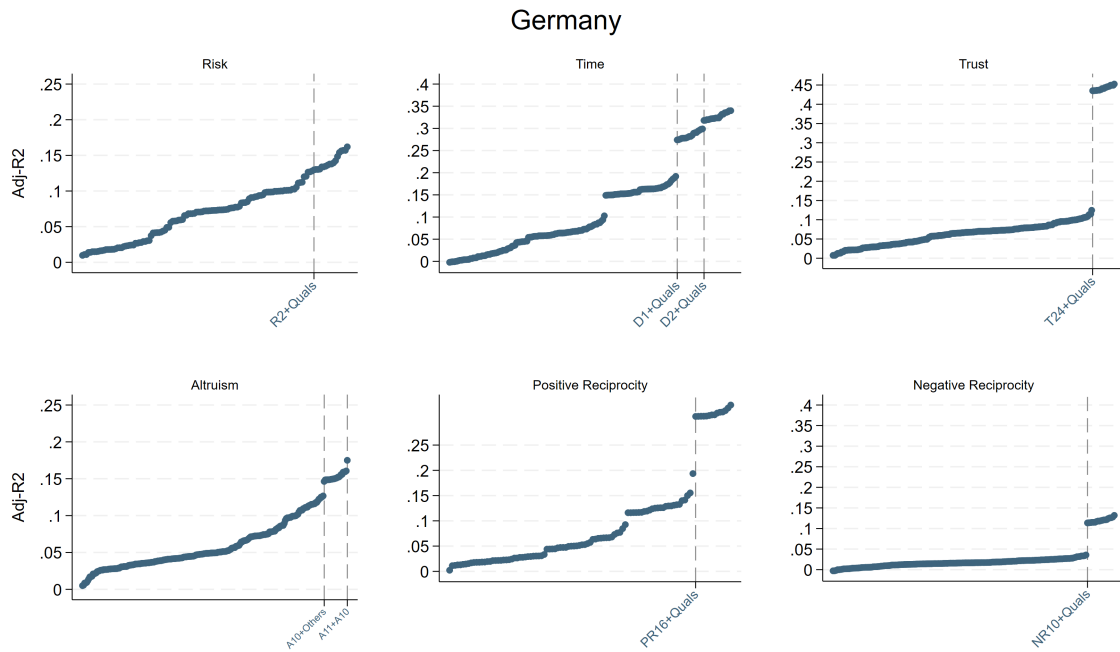


Figure C1: **Sorted adjusted  $R^2$** . This table presents the  $R^2$ s from OLS regressions of the incentivized measure on different combinations of two-items in **Germany**.



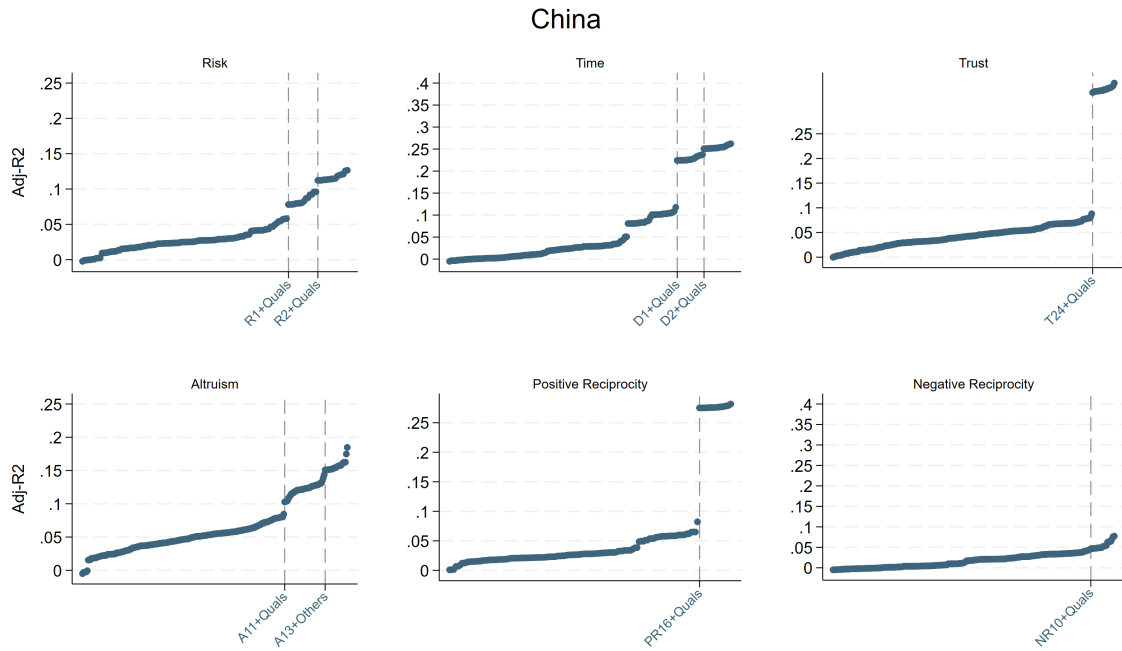


Figure C2: **Sorted adjusted  $R^2$** . This table presents the  $R^2$ s from OLS regressions of the incentivized measure on different combinations of two-items in **China**.

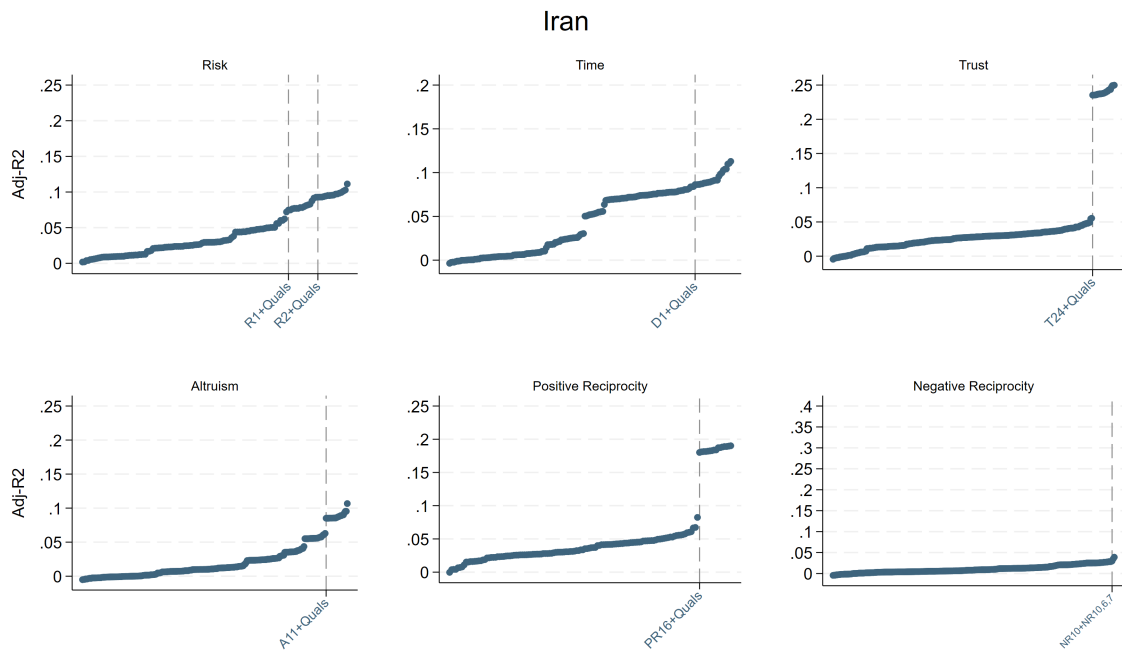


Figure C3: **Sorted adjusted  $R^2$** . This table presents the  $R^2$ s from OLS regressions of the incentivized measure on different combinations of two-items in **Iran**.



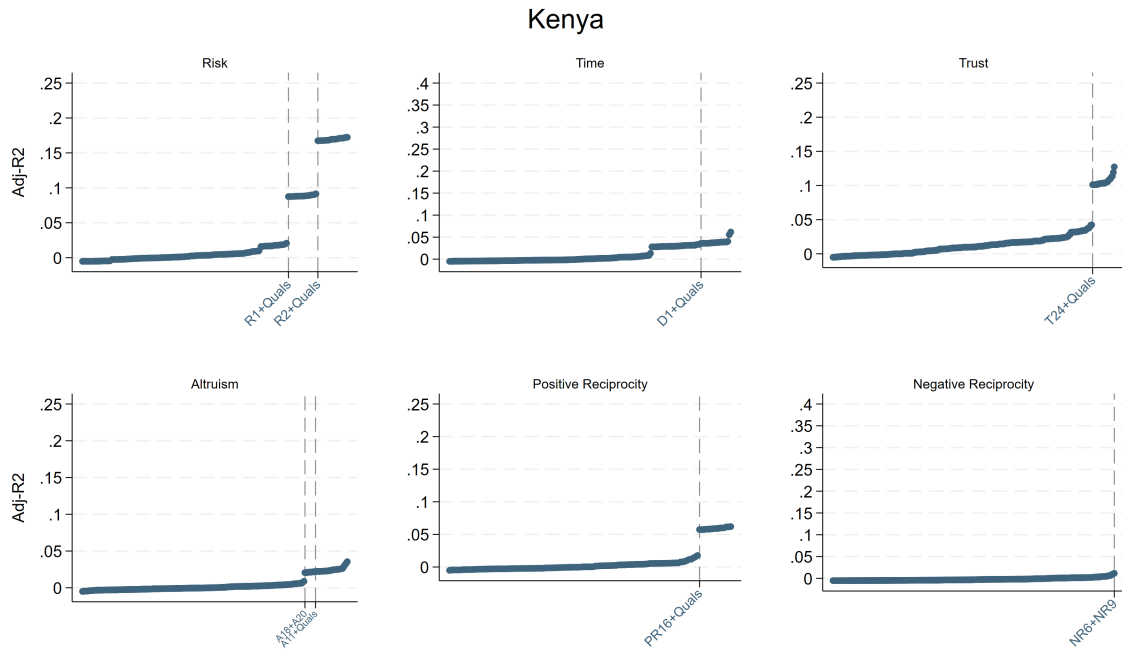


Figure C4: **Sorted adjusted  $R^2$** . This table presents the  $R^2$ s from OLS regressions of the incentivized measure on different combinations of two-items in **Kenya**.

#### C.4 Principal Components and Corresponding Survey Questions for each Preference

See the table on the next page.



Preferences	Principal components	Corresponding survey questions
A: Risk taking	Willingness to take risks Quantitative measures for risk Risk in daily activities Risk in financial investments	R3-R8, R12, R15, R16 Experimental incen. measure, R1,R2 R9-R11 R12-R14, R16
B: Time discounting	Willingness to wait regarding financial decision Quantitative measures for patience Willingness to wait in general Tendency to postpone tasks Compensation for giving up a vacation trip	D5-D10, D15 Experimental incen. measure, D1, D2, D17 D3-D5, D17 D11, D14, D16-D18 D12, D13
C: Trust	Willingness to trust Most people are fair Building trust in other people and their intentions Most people act in their own best interest Quantitative measure for trust Willingness to rely on strangers Willingness to lend to strangers	T1-T8, T13, T14, T17 T20, T21 T10-T13, T16 T22, T23 Experimental incen. measure, T24 T15, T18-T19 T9, T10, T15
D: Altruism	Willingness to share with people Willingness to do good if I expect it returns Quantitative measure for altruism Volunteer to act for good causes I don't understand why some people fight for a cause without direct benefits	A1-A10, A14, A18, A19, A21 A12, A15, A16, A20 Experimental incen. measure, A10, A11, A13 A11, A22, A23 A17, A19, A21-A22
E: Positive reciprocity	Willingness to return a favor I help someone who has helped me before Thank-you gift to return a favor Quantitative measures for positive reciprocity	PR1-PR8 PR11-PR15 PR9, PR10 Experimental incen. measure, PR14, PR16
F: Negative reciprocity	Willingness to punish people who treated others unfairly Willingness to punish people with unfair behavior Dislike being taken for a fool and pushed around Quantitative measures for negative reciprocity Fairness in hypothetical scenario	NR1, NR11-NR14, NR16, NR18 NR2-NR9 NR15, NR17, NR19, NR20, NR22 Experimental incen. measure, NR10 NR2, NR21

Table C3: **Principal components and their corresponding questions for each preference.** This table presents the results of principal component analysis. For each preference, we present components with eigenvalue larger than one and present them in the order of the size of their fraction of explained variance. Components explaining more than 10% of total variance are highlighted in boldface. For each component, we further document the corresponding survey questions.



## **D GPS and PSM Items and Measures**



<b>Panel A:</b>		
<b>GPS</b>	<b>Item description</b>	<b>Weight</b>
Risk taking	R1: Lottery choice sequence using staircase method	0.473
	R3: Self-assessment: willingness to take risks in general	0.527
Patience	D1: Intertemporal choice sequence using staircase method	0.712
	D3: Self-assessment: willingness to wait	0.288
Positive reciprocity	PR9: Gift in exchange for help	0.515
	PR11: Self-assessment: willingness to return a favor	0.485
Negative reciprocity	NR11: Self-assessment: willingness to take revenge	0.374
	NR1: Self-assessment: willingness to punish unfair behavior toward self	0.313
	NR2: Self-assessment: willingness to punish unfair behavior toward others	0.313
Altruism	A11: Donation decision	0.635
	A10: Self-assessment: willingness to give to good causes	0.365
Trust	TR16: Self-assessment: people have only the best intentions	1
<b>Panel B:</b>		
<b>PSM</b>	<b>Item description</b>	<b>Weight</b>
Risk taking	R2: Multiple price list (31 hypothetical choices between a lottery and a safe option).	0.2758
	R3: Self-assessment: willingness to take risks in general	0.2034
Patience	D2: List of 25 hypothetical choices between an early payment “today” and a delayed payment “in 12 months”.	0.4849
	D4: Self-assessment: willingness to wait in comparison with others	−0.1712
Positive Reciprocity	PR11: Hypothetical investment game: second mover behavior	0.4857
	PR9: Gift in exchange for help	0.1640
Negative Reciprocity	NR10: Minimum acceptable offer in hypothetical ultimatum game	0.3284
	NR: Self-assessment: willing to punish unfair behavior	0.1479
Altruism	A11: Donation decision	0.1845
	A10: Self-assessment: willingness to give to good causes	0.3210
Trust	TR24: Hypothetical investment game: first mover behavior.	0.6289
	TR16: Self-assessment: people have only the best intentions	0.1331

Table D1: **Survey Items and Weights of the GPS and the PSM.** This table shows the survey items and weights of the GPS and the PSM. For negative reciprocity, the original self-assessment question on willingness to punish unfair behavior in PSM were split into two questions in GPS.



Preference	Experiment	Measure
Risk taking	Two multiple price lists in which subjects choose between a lottery and varying safe options.	Average of rows in both price lists in which subjects switch from preferring the lottery to the safe option.
Time discounting	Two multiple price lists in which subjects choose between a payment “today” and a larger payment “in 12 months”.	Average of rows in two price lists in which subjects switch from preferring the early to the delayed payment.
Trust	First mover behavior in two investment games.	Average amount sent as a first mover in both investment games.
Altruism	First mover behavior in a dictator game with a charitable organization as recipient.	Amount of donation.
Positive reciprocity	Second mover behavior in two investment games.	Average amount sent back in both investment games.
Negative reciprocity	Investment into punishment after unilateral defection of the opponent in a prisoner’s dilemma (contingent response method) and minimum acceptable offer in an ultimatum game.	Average score: amount invested into punishment and minimum acceptable offer in an ultimatum game.

Table D2: **Overview: Experimental Measures.** This table shows the experimental measures for each preference and experiment.



## **E Experiments and Survey Questions**

### **E.1 Experiments (incl. control questions)**

#### **Risk taking**

We used two multiple price lists to elicit risk preferences. Each list contains 21 decision choices between a lottery and a safe payment. The lottery is always the same and gives 1000 points with a 50% chance, and 0 points with a 50% chance. However, the safe payment varies. In the first list, safe payment increases in steps of 50 points from 0 points to 1000 points. In the second list, safe payment was perturbed before the experiment with points in each choice added or subtracted by a random drawn integer between -5 and +5. All participants faced the same lists and they were informed that in each list, one of the 21 choices would be randomly selected to determine their payoff. The wording of our control questions is as follows:

Before you start making your decisions about the 21 described scenarios, please briefly answer the two questions below. These questions are intended to ensure that you understand the experiment correctly.

Suppose the selected scenario is to choose between the lottery and a sure payment of 450 points.

1. If you decided to choose the lottery, what is the chance that you would get 1000 points?
2. If you decide to choose the sure payment, how many points would you receive?

#### **Time discounting**

We used two multiple price lists to elicit time preference. Each list contains 25 decision choices. In both lists, participants have to choose between receiving a payment of 1600 points today and a higher payment in 12 months. In the first price list, the delayed amount increased in steps of 3.0 percentage points. In the second price list, each delayed payment is perturbed before the experiment by adding or subtracting an amount of up to 0.6 points. All participants faced the same lists and they were informed that in each list, one of the 25 choices would be randomly selected to determine their payoff. They were also informed that in case a delayed payment is chosen to be payoff relevant, they would receive the payment in 12 month, with the amount also adjusted to local inflation rate. Thus, subjects made decisions assuming no inflation. Our control questions are as follows:

Before you start making your decisions about a series of scenarios, please briefly answer the two questions below. These questions are intended to ensure that you understand the experiment correctly.



Suppose that the selected scenario is to choose between receiving 1600.0 points today or 1747.2 points in 12 months.

1. If you decided to choose the payment today, how many points would you receive?
2. If you decided to choose the payment in 12 months, how many points would you receive?

## **Trust**

We used two versions of the investment game introduced by Berg et al. (1995). In the first version, the amount sent by the sender was doubled by the experimenter while in the second version, the amount was tripled. In each version, both the sender and the recipient were endowed with 500 points. We used perfect stranger match so that every participant played both the sender and the recipient in both versions with a different partner. All four decisions were payoff relevant. Our control questions for this game are as follows:

Before you make your decisions, please briefly answer the following questions. As before, these questions are intended to ensure that you understand the experiment correctly.

Suppose you are the recipient. The sender makes a transfer of 150 points. In stage 2, you decide to transfer back 300 points.

1. How many points do you have at the end of this experiment?
2. How many point does the sender have?

Now, suppose you are the sender. You decide to transfer 0 points to the recipient. In stage 2, the recipient decides to transfer back 0 points.

1. How many points do you have at the end of the experiment?
2. How many points does the recipient have?

## **Altruism**

We used a dictator game with charitable organizations as the recipient. Participants were endowed with 300 points and had to allocate the points between themselves, and one charitable organization. In the original German version, the charitable organizations are: Brot für die Welt, Kindernothilfe, German Red Cross,



Welthungerhilfe, Bund für Umwelt und Naturschutz Deutschland, Greenpeace, Terre des Hommes, and Aktion Mensch. Participants were also given the opportunity to nominate a charitable organization for receiving the money, if they did not want to choose one from the given list. To adapt the charitable organizations to other countries, we kept international organizations such as Red Cross and Greenpeace, but replaced German charitable organizations with domestic charitable organization in China, Iran and Kenya. Since donations were done by the experimenter on behalf of participants, we made it clear that all participants can access to a cloud folder in which donation receipts are stored and participants can verify their own donation. Our control questions are as follows:

Before you make your decision, please briefly answer the following two question. These questions are intended to ensure that you understand the experiment correctly.

Suppose you receive from us an amount of 300 points, which you can use to make a donation.

You decide to donate 300 points. How many points would you receive from this experiment?

Suppose you decide to donate 0 points. How many points would you receive from this experiment?

### **Positive reciprocity**

We elicited positive reciprocity based on the second mover behavior in the above-mentioned investment games. We used the strategy method to elicit the second-mover behavior and took the average of these decisions in each version of the game. The average from the two games is then used to measure positive reciprocal inclination.

### **Negative reciprocity**

We measure negative reciprocity with two experiments. The first is the ultimatum game introduced by Güth et al. (1982). Participants played twice and were randomly assigned to be the sender in one game, and the recipient in the other game. In each game, the sender had to decide how to split 500 points between themselves and the recipient. We take the recipient's minimum acceptable offer as the first measure for negative reciprocity. Our control questions for the ultimatum game are:

Before you make your decisions, please briefly answer the following questions. These questions are intended to ensure that you understand the experiment correctly.

Suppose you are assigned the role of the sender. What is the maximum number of points you may send



to the recipient?

Now, suppose you send 50 points but the recipient wants to receive at least 100 points so that s/he is willing to accept the proposal.

1. How many points do you earn at the end of the experiment?
2. How many points does the recipient earn?

The second experiment is prisoner's dilemma with a subsequent punishment stage (Falk et al., 2005). The experiment is framed as a project in which both participants could decide to contribute or not in stage 1. In stage 2, participants can deduct points from the other participant at the cost of 1/3 of a point. In other words, it costs self one point to deduct three points from the other person. Participants were endowed with 300 points. When both contribute, both receive 480 points. When neither contributes, both keep their 300 points. When one contributes whereas the other one does not, then the contributed one receives 240 points whereas the person who did not contribute receives 540 points. We use the amount participants invested into punishment in case of unilateral defection of the other player as the second measure for negative reciprocity. Overall then, negative reciprocity is the average of both standardized measures.

For the prisoner's dilemma game, our control question for stage 1 is as follows:

Before continuing to the experiment, please briefly answer the following question.

True or false? If you contribute, but the other participant does not contribute, you get more points than the other participant.

In stage 2, the control questions are:

Before you make your decisions, please briefly answer the following questions. As before, these questions are intended to ensure that you understand the experiment correctly.

Suppose both you and the other participant decide to not contribute, and suppose the other participant deducts 60 points from you whereas you do not deduct any points from him/her.

1. How many points do you earn at the end of the experiment?
2. How many points does the other participant earn?



## E.2 Survey items

During the course of conducting our study, we note that the survey batteries utilized in the original study can be shortened as they contain questions with similar content, questions that are unlikely to capture individual variance, require respondents to possess a specific characteristic (e.g., employment), refer to a context that is illegal or contrary to social norms in some countries (e.g., gambling), or do not specify to whom the answer/reaction is directed (e.g., who receives a gift, who is the transgressor of a norm, who gets punished, etc.). Additionally, some questions measure two distinct preferences simultaneously (e.g., both positive and negative reciprocity), complicating the interpretation of the results. Many of these questions in survey batteries lacked explanatory power for predicting laboratory behavior, neither in Bonn nor in Tehran. Under this ex-post knowledge, we wonder whether these queries are necessary to be included in the battery. Mainly, excluding these questions makes the survey part much shorter and helps remain the focus of participants, which is important in reducing the measurement error. As a result, we view these as opportunities to enhance the original survey battery in order to expedite implementation and increase efficiency. We used an iterative technique to include the most pertinent questions for each preference.

First, we incorporate all survey items which are included in the GPS (Falk et al. (2018)). Second, we include all survey items which are not in GPS but in the PSM (Falk et al. (2023)). Third, we add survey items that are neither GPS nor PSM but were found to be highly correlated with behavior in either the Bonn or Tehran incentivized lab studies (Falk et al., 2023; Kosfeld and Sharafi, 2023). We have reviewed both selected and remaining items up to this point. When an item is selected from a set of questions covering various contexts, we also include the remaining questions in this set to complete the context. This also applies to the set of triangulation questions for each preference: self-assessment, self-assessment in contrast to others, and self-assessment in relation to how other people judge them. We then further remove some questions because they referred to activities that are uncommon or illegal in some countries. Finally, we conduct a re-evaluation of all selected items. After the above mentioned iteration, we notice that the original survey batteries for trust and altruism omit only one or two questions. Due to the low expense of adding this one or two questions, we also decided to include them. This means that we retained the entire battery of survey questions for trust and altruism, whereas, for other preferences, we retained only a subset of the most important survey questions.

Concerning the wording of all survey questions, for GPS questions, we use the exact English wording



as well as the provided translations in Chinese and Farsi provided by Falk et al. (2018). The remaining survey questions are worded exactly as Falk et al. (2023) presented in their online appendix. However, in some cases, we made modest adjustments. These modest improvements, we believe, are necessary and more accurate. This Appendix contains the final, and comprehensive list of survey items used in our study in the English wording.

## **Risk taking**

- **(R1/GPS) Switching row from staircase measure.** Please imagine the following situation. You can choose between a sure payment of a particular amount of money, OR a draw, where you would have an equal chance of getting 300 Euro or getting nothing. We will present to you five different situations. E.g.

1. What would you prefer: a draw with a 50- percent chance of receiving 300 Euro, and the same 50-percent chance of receiving nothing, OR the amount of 160 Euro as a sure payment?

(a) 50/50 chance – go to question 17

(b) Sure payment – go to question 2

- **(R2) Hypothetical list lottery vs. increasing safe option.** In this part of the survey you can choose between a sure payment and a lottery. This lottery is the same as before. With a 50% chance you get 300 Euro and with 50% chance you get 0 Euro. The sure payment is different in each scenario.

Scenario 1. Lottery: 50% chance of 300 Euro and 50 % chance of 0 Euro, or Sure payment: 0 Euro

...

Scenario 31. Lottery: 50% chance of 300 Euro and 50 % chance of 0 Euro, or Sure payment: 300 Euro

- **(R3/GPS) Willingness to take risks.**

How do you assess yourself? Please tell me, in general, how willing or unwilling you are to take risks, using a scale from 0 to 10, where 0 means you are “completely unwilling to take risks” and 10 means you are “very willing to take risks”. You can also use any number between 0 and 10 to indicate where you fall on the scale, using 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.

Completely unwilling to take risks=0 ... 10=Very willing to take risks



- **(R4) Willingness to take risks compared with others.**

How do you assess yourself in comparison to others? Are you a person who is willing to take risks or do you try to avoid risks?

- **(R5) Willingness to take risks assessed by others.**

- It's not always easy to assess yourself well. How sure are you that your assessment of the last three questions is accurate?

- **(R6) Willingness to take risks in financial investments.**

One can behave differently in various situations. How do you assess your willingness to take risks in the following contexts?

Please click a button on the scale, where 0 represents "not willing to take risks at all" and 10 represents "very willing to take risks". With the values in between you can grade your assessment. Not willing to take risks at all=0 o o o o o o o o o 10= Very willing to take risks

- **(R7) Willingness to take risks in important decisions in life.**

- **(R8) Willingness to take risks in professional life.**

- **(R9) Willingness to take risks in leisure/sports.**

- **(R10) Willingness to take risks in traffic.**

- **(R11) Willingness to take risks in dealing with other people.**

- **(R12) How likely: you invest 5% of your annual income in a very speculative stock**

Please click a button on the scale, where 0 means "Very unlikely" and 10 means "Very likely". With the values in between you can grade your assessment. How likely is it, that: You invest 5% of your annual income in a very speculative stock? Very unlikely=0 ... 10= Very likely

- **(R13) How likely is it that: I often behave according to the motto: It is better to be safe than sorry.**

- **(R14) How likely is it that: I avoid risky things.**



- **(R15) How likely is it that: I like taking risks.** (Instrument for R3)
- **(R16) You win 100,000 EUR in a lottery. How much of the money you won in the lottery would you invest in the risky yet profitable lottery?**

How would you decide? Please think about what you would do in the following situation. Imagine you win 100.000 Euro in a lottery. Immediately after receiving the money you get an offer to participate in the following lottery: There is a chance to double the money. But there is an equally high chance to lose half of the money invested in the lottery. You can participate in the lottery using the whole amount you won or only a part of it. We would like to know: How much of the money you won in the lottery would you invest in the risky yet profitable lottery?

1. The whole amount of 100.000 Euro
2. 80.000 Euro
3. 60.000 Euro
4. 40.000 Euro
5. 20.000 Euro
6. Nothing at all

- Please indicate how much we can rely on your answers from the questions in the previous survey. Please click a button on the scale from 0 to 10.

You cannot rely on my answers=0 ... 10=You can rely on my answers

## **Time discounting**

- **(D1/GPS) Switching row from staircase procedure.** Suppose you were given the choice between receiving a payment today or a payment in 12 months. We will now present to you five situations. The payment today is the same in each of these situations. The payment in 12 months is different in every situation. For each of these situations we would like to know which you would choose. Please assume there is no inflation, i.e., future prices are the same as today's prices. E.g.

1. Please consider the following: Would you rather receive 100 Euro today or 154 Euro in 12 months?



- (a) Today – go to question 17
- (b) In 12 months – go to question 2

- **(D2) Switching row from hypothetical list.**

We will present you with a series of hypothetical scenarios in which you need to choose between a payment today and a payment in 12 months. The payment today stays the same but the payment in 12 months changes with every scenario.

Scenario 1: Payment today 100.0 Euro, Payment in 12 months 100.0 Euro ... Scenario 25: Payment today 100.0 Euro, Payment in 12 months 185.0 Euro

- **(D3/GPS) Willingness to wait.** How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?

Completely unwilling to do so= 0 ... 10=Very willing to do so

- **(D4) Willingness to wait compared with others.**

- **(D5) Willingness to wait assessed by others.**

- **(D6) Willingness to wait wrt. financial investments.** How would you assess your willingness to give up something today in order to benefit from that in the future in the following contexts? One can behave differently in various situations. How would you assess your willingness to give up something today in order to benefit from that in the future in the following contexts?

- **(D7) Willingness to wait wrt. important decisions in life.**

- **(D8) Willingness to wait wrt. professional career.**

- **(D9) Willingness to wait wrt. bigger purchases.**

- **(D10) Willingness to wait wrt. a longer journey/trip.**

- **(D11/GPS) I tend to postpone tasks even if I know it would be better to do them right away.**

- **(D12) Won vacation.** Imagine the following situation:



Imagine you had won a 10-day trip for two people worth 2,000 Euro to an exciting destination. Due to high demand you are asked whether you would be willing to wait three years before making the trip. In return for waiting you would be given an extension of the trip.

Please consider: how many extra days would one have to offer you for you to be willing to postpone the trip for three years?

- **(D13) Won vacation.**

If it was possible to exchange the trip for money: how much money would one need to offer you for you to be willing to forgo the trip?

- **(D14) I try hard to always have some extra money for unexpected expenditures.** The following statements characterize different attitudes towards life and the future.

Please tell us to what extent you agree with each statement. Please indicate your answer on a scale from 0 to 10. A 0 means “does not apply at all” and a 10 means “applies fully”.

I try hard to always have some extra money for unexpected expenditures.

- **(D15) I give up something today so that I can afford more tomorrow.** (Instrument for D3)
- **(D16) I would rather have some fun today and not think about tomorrow.**
- **(D17) My monthly expenses often exceed what I can afford.**
- **(D18) I am a person who often does not keep their own good resolutions.**
- **(D19) How much money do you save per month?**

Please think: How much money do you save per month? Please try to specify the amount you save per month as exactly as possible.

## **Trust**

- **(TR1) Willingness to trust.** (Instrument for TR16) How do you assess yourself? Generally speaking, are you a person who is willing to trust other people, or are you not willing to trust other people? Please click a button on the scale, where 0 represents “not willing to trust at all” and 10 represents “very willing to trust”. With the values in between you can grade your assessment.



not willing to trust at all = 0 ... 10 = very willing to trust

- **(TR2) Willingness to trust compared to others.**
- **(TR3) Willingness to trust assessed by others.**
- **(TR4) Willingness to trust people in your hometown.**
- **(TR5) Willingness to trust your circle of friends.**
- **(TR6) Willingness to trust people in your professional environment.**
- **(TR7) Willingness to trust strangers.**
- **(TR8) Willingness to trust neighbors.**
- **(TR9) Willingness to lend money to a stranger.** What would you choose? Imagine the following situation. You are on vacation in a foreign country. A person, whom you meet in your hotel but whom you do not know, asks you for a favor. He or she urgently needs cash in order to pay for their partner's doctor visit and promises to pay you back the following day. How much money would you be willing to lend to that person?
- **(TR10) How often give a hitchhiker a ride.**
- **(TR11) How often leave your personal belongings unattended in a public place.**
- **(TR12) How often do not lock the apartment door.**
- **(TR13) In comparison to others I quickly build trust in strangers.**
- **(TR14) Other people regard me as too credulous and trusting.**
- **(TR15) I find it difficult to talk about personal issues with people I have not known for a long time yet.**
- **(TR16/GPS) I assume that people have only the best intentions.**
- **(TR17) In general, one can trust other people.**
- **(TR18) Nowadays one cannot rely on anyone anymore.**



- **(TR19) When dealing with strangers it is better to be careful before one relies on them.**
- **(TR20) Most people would take advantage of you when they have the chance.** What do you think, which statement applies more? Do you think... that most people would take advantage of you when they have the chance, or...
- **(TR21) ... Most people would be fair to you.**
- **(TR22) Most people try to be helpful.**
- **(TR23) Most people only act in their own best interest.**
- **(TR24) 1st mover decision in hypothetical trust game.** How would you decide?

Please consider what you would do in the following situation: You and a person whom you do not know both have to make a decision about the allocation of money and together you achieve an outcome. The rules are the following: both of you get an account with 20 Euro. Thus, at first, both you and the other person have 20 Euro each on their account. The other person has to decide first. She can transfer money to your account. She can transfer any round amount, i.e. 0 Euro, 1 Euro, 2 Euro, etc. up to 20 Euro. Each Euro that the other person decides to transfer to you is tripled by the people conducting the study and then credited to your account. Thus, after the first step, the other person has 20 Euro minus the amount she transferred to you on her account. You, on the other hand, have 20 Euro plus three times the amount that was transferred to you on your account. Now you have to make a decision. You can transfer money back to the other person. You can transfer any amount to the other person, i.e. 0 Euro, 1 Euro, 2 Euro, etc. up to 80 Euro depending on how much money is on your account after receiving the transfer from the other person. After this decision, the study is over, and the amounts on the two accounts are final. The other person has 20 Euro minus the amount she transferred to you plus the amount you transferred back on her account. You have 20 Euro plus three times the amount the other person transferred to you minus the amount you transferred to the other person on your account. For a given transfer of the other person we would now like to know how much money you would decide to transfer back. Assume you were in the position of the other person and had to decide which amount to transfer. Which amount would you transfer?



## Altruism

- **(A1) Willingness to share.** How do you assess yourself? Are you a person who is generally willing to share with others without expecting something in return, or are you not willing to do so? Please click a button on the scale, where 0 represents “not willing to share without expecting something in return” and 10 represents “very willing to share even without expecting something to return“. With the values in between you can grade your assessment.

not willing to do so =0 ... 10= very willing to do so

- **(A2) Willingness to share compared with others.**
- **(A3) Willingness to share assessed by others.**
- **(A4) Willingness to share with people in hometown.** How do you assess your willingness to share with others without expecting anything in return in the following contexts: With people in your hometown. One can behave differently in various situations. Please again indicate your answer on a scale from 0 to 10. A 0 means “completely unwilling to do so,” and a 10 means “very willing to do so.” You can also use any number between 0 and 10 to indicate where you fall on the scale, using 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.
- **(A5) Willingness to share with friends.**
- **(A6) Willingness to share with people from your professional environment.**
- **(A7) Willingness to share with strangers.**
- **(A8) Willingness to share with neighbors.**
- **(A9) Willingness to share with people in need.**
- **(A10/GPS) Willingness to give to good causes.**
- **(A11/GPS) You unexpectedly received 1000 EURO, how much would you donate to a good cause.**

Imagine the following situation: Today you unexpectedly received 1,000 Euro. How much of this amount would you donate to a good cause? (values between 0 and 1000 are allowed)



- **(A12) At work, I am only willing to do something for a colleague if I expect that they would do the same for me.** How well do the following statements describe you as a person? Please click a button on the scale, where 0 represents “Does not describe me at all” and 10 represents “Describes me perfectly”. With the values in between you can grade your assessment.

At work, I am only willing to do something for a colleague if I expect that they would do the same for me.

- **(A13) I am willing to donate time and money to charity, even if I do not profit from that directly.**  
(Instrument for A10)
- **(A14) I am willing to help others even if I expect that I will never meet them again.**
- **(A15) When I spend time and money on something I expect to profit from that in the future.**
- **(A16) When I donate money I expect that this is recognized and acknowledged.**
- **(A17) I do not understand why some people spend their lifetime fighting for a cause which they do not benefit from directly.**
- **(A18) I am a person who would give their shirt off their back to help others.**
- **(A19) In comparison to others I am a rather selfless person.**
- **(A20) I am only willing to help others if I expect that they would do the same for me.**
- **(A21) Other people regard me as an unselfish person.**
- **(A22) How many hours per month do you volunteer for good causes.**
- **(A23) How many people know that you commit time to charitable purposes.**

### **Positive reciprocity**

- **(PR1/GPS) Willingness to return a favor.** How do you assess yourself? Are you a person who is generally willing to go out of their way to return a favor or help even if it is costly, or are you not willing to do so? Please click a button on the scale, where 0 represents “not willing to incur costs to



return a favor” and 10 represents “very willing to incur costs to return a favor”. With the values in between you can grade your assessment.

not willing to do so =0 ... 10= very willing to do so

- **(PR2) Willingness to return a favor compared with others.**
- **(PR3) Willingness to return a favor assessed by others.**
- **(PR4) Willingness to return a favor to people in your hometown.** How do you assess your willingness to return a favor or help in the following contexts? One can behave differently in various situations. Please click a button on the scale where 0 means “not willing to incur costs to return a favor” and 10 means “very willing to incur costs to return a favor”. With the values in between, you can grade your assessment.

When it comes to people in your hometown.

- **(PR5) Willingness to return a favor to circle of friends.**
- **(PR6) Willingness to return a favor to people in your professional. environment**
- **(PR7) willingness to return a favor to strangers.**
- **(PR8) Willingness to return a favor to neighbors.**
- **(PR9/GPS) Lost-in-the-city: which present do you give to the stranger.** How would you decide? Helping you costs the stranger about 20 Euro in total. However, the stranger says he or she does not want any money from you. You have six presents with you. The cheapest present costs 5 Euro, the most expensive one costs 30 Euro. Do you give one of the presents to the stranger as a “thank you” gift?
- **(PR10) Medical-help: how much do you spend on a thank-you present.** How would you decide? Assume that you are abroad and need medical treatment. In the country you are in it is common that the doctor treats patients only for cash. The treatment costs about 100 Euro. You do not have any cash with you. A stranger in the waiting room observes the situation and gives 100 Euro as a gift to you. You are happy to take the gift. You ask the stranger for their address. When returning home two



weeks later you decide that you want to thank the stranger and send them a present. How much do you spend on a present that you then send to the stranger?

- **(PR11/GPS) When someone does me a favor, I am willing to return it.**
- **(PR12) I go out of my way to help someone who has helped me before.**
- **(PR13) I am willing to incur costs to help someone who has helped me before.**
- **(PR14) I do not like the feeling of owing something to someone.**
- **(PR15) If a colleague does me a favor at work, I make sure to return the favor at the next occasion, even if I have to invest precious time to do so.**
- **(PR16) 2nd mover behavior in hypothetical trust game.** a) Assume that the other person transfers 5 Euro to your account. After the first step you have  $20+3*5$  Euro = 35 Euro, the other person has  $20-5$  Euro = 15 Euro. Which amount do you transfer back? b) Assume that the other person transfers 10 Euro to your account. After the first step you have  $20+3*10$  Euro = 50 Euro, the other person has  $20-10$  Euro = 10 Euro. Which amount do you transfer back? c) Assume that the other person transfers 15 Euro to your account. After the first step you have  $20+3*15$  Euro = 65 Euro, the other person has  $20-15$  Euro = 5 Euro. Which amount do you transfer back? d) Assume that the other person transfers 20 Euro to your account. After the first step, you have  $20+3*20$  Euro = 80 Euro, the other person has  $20-20$  Euro = 0 Euro. Which amount do you transfer back?

### **Negative reciprocity**

- **(NR1/GPS) willingness to punish people who treat you unfairly.** How do you assess yourself? Please again indicate your answer on a scale from 0 to 10. A 0 means “completely unwilling to do so,” and a 10 means “very willing to do so.” You can also use any number between 0 and 10 to indicate where you fall on the scale, using 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10. How willing are you to punish someone who treats you unfairly, even if there may be costs for you?  
  
completely unwilling to do so =0 ... 10= very willing to do so
- **(NR2/GPS) Willingness to punish people who treat others unfairly.**



- **(NR3) Willingness to punish compared with others.**
- **(NR4) Willingness to punish assessed by others.**
- **(NR5) Willingness to punish people in your hometown.**
- **(NR6) Willingness to punish your circle of friends.**
- **(NR7) Willingness to punish people in your professional environment.**
- **(NR8) Willingness to punish strangers.**
- **(NR9) Willingness to punish neighbors.**
- **(NR10) Minimum acceptable offer in hypothetical Ultimatum Game.** How would you decide?  
Imagine the following situation: Together with a person whom you do not know you won 100 Euro in a lottery. The rules stipulate the following: One of you has to make a proposal about how to divide the 100 Euro between you two. The other one gets to know the proposal and has to decide between two options. He or she can accept the proposal or reject it. If he or she accepts the proposal, the money is divided according to the proposal. If he or she rejects the proposal, both receive nothing. Assume that the other person makes the proposal about how to divide the money. You, on the other hand, have to decide whether to accept or reject the proposal. What is the minimum amount the other person has to offer you for you to be willing to accept the proposal?
- **(NR11/GPS) If I am treated very unjustly, I will take revenge at the first occasion, even if there is a cost to do so.**
- **(NR12) When someone puts me into a difficult situation I will do the same to them.**
- **(NR13) If someone insults me I will also behave in an insulting way towards them.**
- **(NR14) If someone harms me on purpose I will try to give that person a taste of their own medicine.**
- **(NR15) I am not a person who is taken for a fool.**
- **(NR16) If someone behaves unfairly towards me in sports, I will also behave unfairly towards them.**



- **(NR17) I am not a person who lets others push me around.**
- **(NR18) When someone treats me in a bad way, I do not just let it go.**
- **(NR19) I absolutely dislike being the fool.**
- **(NR20) You sometimes have to play tough in order not to be taken advantage of.**
- **(NR21) hypothetical scenario: cutting wages fairness** Imagine the following scenario: A business in a city with a high level of unemployment makes profits despite a recession. The enterprise's chairman announces a decision to cut all wages and salaries by 5%. How fair do you think is this decision? Please click a button on the scale, where 0 represents "absolutely unfair" and 10 represents "absolutely fair". With the values in between you can grade your assessment.  
  
absolutely unfair = 0 ... 10 = absolutely fair
- **(NR22) It is important to me to be respected by others.**

### **Real-life behavior**

- **(DIM1) How tall are you?** How tall are you in cm? (if you don't know, please estimate)
- **(DIM2) How many kilos do you weigh?** In kgs.
- **(DIM3) Do you currently smoke?** Do you currently smoke, be it cigarettes, a pipe, or cigars? Yes or No.
- **(DIM4) Do you exercise?** Do you exercise or do sports regularly? Yes or No.
- **(DIM5) Do you save money?** Do you save money? Yes or No.
- **(DIM6) Do you invest in securities?** Do you invest in securities (e.g., stocks, funds, bonds, equity warrants)? Yes or No.
- **(DIM7) Do you participate in any social organization?** Do you participate in any social organization (e.g., charity, community or neighborhood committee, religious group, cultural or sports group, education-oriented programs or groups, environment conservation, labor union, political party)?
- **(DIM8) What are the average hours in a month you spend in social organizations?**



- **(DIM9) Do you regularly use reusable bags when shopping?**
- **(DIM10) Do you regularly conserve water when it is not directly needed?**
- **(DIM11) Do you regularly use a reusable cup/container for drinking?**
- **(DIM12) How many close friends do you have?**
- **(DIM13) Do you plan to pursue a higher level of education or go to the job market directly after your studies?**
- **(DIM14) Do you spend more time preparing for exams in advance or rather only prepare shortly before the exam?**
- **(DIM15) Do you prefer team-oriented sports or individual-oriented sports?**
- **(DIM16) On average, what is the relationship duration between you and your close friends?**
- **(CH1) Age in 2023.**
- **(CH2) Gender.**
- **(CH3) In what type of program are you now registered in?** In what type of program are you now registered? Bachelor's, Master's, Doctorate, Other.
- **(CH4) In which year of your studies are you now?**
- **(CH5) Your academic performance in the university.** Which one best describes your academic performance in the university? Very low, Low, Medium, High, Very high, Prefer not to say.
- **(CH6) The highest educational attainment achieved by either of your parents (or legal guardians).** Primary school, Middle school, High school, Trade/vocational diploma or certificate, College diploma or Certificate, Bachelor, Master, Higher, Prefer not to say.
- **(CH7) Your family income class.** Which one best describes your family income class? Upper (Wealthy) class, Upper-middle class, Middle-class, Working-class, Low-income, Prefer not to say.
- **(CH8) Do you have prior knowledge about the experiments?**
- **(CH9) Economics.**



- **(CH10) Education.**
- **(CH11) Engineering.**
- **(CH12) Humanities.**
- **(CH13) Journalism, Media, communication.**
- **(CH14) Management, business.**
- **(CH15) Math, statistics, informatics.**
- **(CH16) Medicine.**
- **(CH17) Natural sciences.**
- **(CH18) Social sciences.**
- **(CH19) Law.**
- **(CH20) Others.**
- **(UQ1/GPS) I am good at math.** Please indicate your answer on a scale from 0 to 10 A 0 means “does not describe me at all,” and a 10 means “describes me perfectly.” You can use any number between 0 and 10 to indicate where you fall on the scale, using 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10. Does not describe me at all = 0 ... 10 = Describes me perfectly.

### **E.3 Excluded questions from the original survey battery in Falk et al. (2023)**

#### **Risk taking**

- How likely is it, that...
  - that you admit that your tastes are different from those of your friends?
  - you go camping in the wild, far away from civilization or campgrounds?
  - you buy an illegal drug for your own use?
  - you invest 10% of your annual income into investment funds with moderate growth rates?
  - you drink five or more alcoholic drinks on one evening?



- you cheat substantially on your income tax?
- you disagree with your father on a major issue?
- you have an affair with a married man or woman?
- you forge somebody's signature?
- you present somebody else's work as your own?
- you go on vacation in a third-world country without a pre-arranged travel route and without booking accommodations ahead?
- you argue with a friend who has a very different opinion on an issue?
- you ask your boss for a raise?
- that you illegally copy a piece of software?
- you go whitewater rafting at high water in the spring?
- you tell a friend that his/her partner flirted with you?
- you invest 5% of your annual income in a conservative stock?
- you shoplift a small item (e.g., a pen or a lipstick)?
- you wear unconventional or provocative clothes?
- you engage in unprotected sex?
- you steal an additional TV cable connection?
- you don't wear a seatbelt when in the front seat?
- you invest 10% of your annual income in government bonds (treasury bills)?
- you periodically engage in a dangerous sport (e.g., mountain climbing or sky diving)?
- You gamble away a week's income at a casino.
- You take a job that you like instead of a job that is very reputable but that you like less.
- You openly express an opinion or viewpoint that is unpopular but of which you are convinced.
- You don't wear sunscreen when you expose yourself to the sun.
- You try bungee jumping at least once in your life.
- You fly a small plane if you could.



- You walk alone through a rather unsafe part of the city at night.
- You regularly eat high-cholesterol food.
- Imagine you won a prize in a lottery. You can choose between two payment options. Either you get a raffle ticket or you get a safe payment. If you decide to take the raffle ticket, you receive 1,000 Euro with a probability of 50% and you receive nothing with a probability of 50%. Please consider: How much money would the safe payment need to be in order for you to prefer it over the raffle ticket?
- Imagine the following situation: you are the only member of your household that has a monthly income, and you have a good job which would guarantee your family income for the rest of your life. Now you have the option to take a new and equally good job. The payment at this new job is variable, so that your household income will double with a probability of 50% and will decrease by 30% with the same probability. Would you be willing to take the new job?

### **Time discounting**

- How well do the following statements describe you as a person?
  - I often realize that I make decisions knowing that I will regret them in the future.
  - I often think about the future.
  - I find it hard to resist unhealthy but delicious food.
  - I am a person who does not care about tomorrow and who only lives for the moment.
  - I am a person who often regrets my own decisions.
  - I am a person who often acts hastily/prematurely.
  - I save for my retirement.
  - I do not find it hard to resist temptations.
  - I spend too much money.
  - I eat too much.
  - I work out too little.
  - I wish I was more self-disciplined.



- Usually, I am sufficiently prepared for exams.
  - I often act without considering all alternatives.
  - I tend to interrupt people in conversations.
  - Once I set a goal for myself, I usually achieve it.
- How well do the following statements apply to you:
    - I have a good idea of what my next job will look like.
    - My life at the moment is completely different from what I imagined it would be like three years ago.
    - I have a precise idea/clear picture of what I can expect in the upcoming year.
    - Last year went very differently from what I previously expected.
  - When I have to make an important decision, I try to paint a clear picture/get a precise idea of the consequences of that decision.
  - When I make an important decision, the outcome usually corresponds with what I have imagined it to be.
  - The following statements characterize different attitudes towards life and the future.
    - I try hard to always have some extra money for unexpected expenditures.
  - You suddenly got into an unforeseen situation, and you had to pay about 1,000 Euro within two weeks: could you manage that?

### **Positive reciprocity**

- Imagine the following situation: you are shopping in an unfamiliar city and realize you lost your way. You ask a stranger for directions. The stranger offers to take you with their car to your destination. The ride takes about 20 minutes and costs the stranger about 20 Euro in total. The stranger does not want money for it. You have six bottles of wine with you. One bottle costs 5 Euro. You decide to give a bottle to the stranger as a thank-you gift. How many bottles do you give? (Options: One/two/three/four/five/six bottles.)



## **Negative reciprocity**

- Please consider what you would do in the following situation: you and a stranger are involved in a car accident. You are not to blame for the accident, but the stranger claims that you ran a red light even though it was the stranger himself who ran the red light. Even though the stranger's claim is false, the claim is believed to be correct and you have to pay a fine of 300 Euro. There was an eyewitness who saw what really happened. If the eyewitness testifies, you don't have to pay the fine but the stranger has to instead. In addition, the stranger will then have to pay a fine for making a false testimony. Assume that there is a detective who will definitely find the eyewitness, and that the eyewitness will testify if the detective finds him. What is the maximum amount of money that you are willing to spend on hiring the detective?
- Imagine the following scenario: It is the weekend of the annual fair, which is well-attended as usual. It is warmer than expected, so that the people at the fair drink much more than in the preceding years. As a result, the hosts decide to raise the prices of the drinks. How fair do you think is this decision?
- Imagine the following scenario: The preparation of the annual accounts is coming up for the business you are employed by. Hence, all employees have to work overtime in order to manage and finish the workload that the boss expects from them. Nevertheless, one of your co-workers leaves the office every day at the usual time, so that you and the other colleagues additionally have to take on his workload as well. Please express the intensity of your feelings towards that co-worker.
  - How upset are you on a scale from 0 to 10?
  - How angry are you on a scale from 0 to 10?

## **Questions addressing both positive and negative reciprocity**

- Are you a person who is generally willing to reward fair behavior and punish unfair behavior even if it is costly, or are you not willing to do so?
- In comparison to others, are you a person who is generally willing to reward fair behavior and punish unfair behavior, even if it is costly, or are you not willing to do so (in comparison to others)?



- Do other people assess you as a person who is generally willing to reward fair behavior and punish unfair behavior even if it is costly, or as a person who is not willing to do so?