

Measuring Economic Preferences with Behavioral Experiments and Surveys Across the Globe

Michael Kosfeld¹, Zahra Sharafi², Maíra Sontag González³, and Na Zou⁴

¹Goethe University Frankfurt, CEPR, CESifo, IZA, SAFE, and ZEF

²Frankfurt School of Finance and Management

³Goethe University Frankfurt

⁴University of Bath

January 6, 2026

Abstract

Recent work in economics has developed survey-based measures of economic preferences that are “experimentally” or “behaviorally validated.” We scrutinize this approach focusing on the Global Preference Survey (Falk et al., 2018, 2023). Replicating their behavioral validation on almost 2,000 participants in five countries, China, Colombia, Iran, Kenya, and the US, we find that many items correlate significantly with behavior in incentivized choice experiments, but coefficients vary and are not always sizable. While hypothetical versions of an experiment consistently show the strongest correlations, qualitative self-assessments are also almost always among the best single-item modules predicting behavior in all countries. Within- and across-preference principal component analyses, as well as comparison in terms of predictive validity, suggest that qualitative self-assessments and behavioral measures capture distinct dimensions of the same latent construct of an economic preference, casting doubt on the use of one instrument as an easy substitute for the other.

Corresponding author: Zahra Sharafi, Frankfurt School of Finance and Management, Adickesallee 32-34, 60322 Frankfurt am Main, Germany. Email: z.sharafi@fs.de. Michael Kosfeld: Faculty of Economics and Business, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt am Main, Germany. Email: kosfeld@econ.uni-frankfurt.de. Maíra Sontag González: Faculty of Economics and Business, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt am Main, Germany. Email: sontaggonzalez@econ.uni-frankfurt.de. Na Zou: University of Bath School of Management, Claverton Down, Bath, BA2 7AZ. Email: nz464@bath.ac.uk. We thank Sen Geng, Johannes Haushofer, Rosa Hendijani, the Busara Center for Behavioral Economics, Inc., Ryan Oprea, and Julia Seither Afonso for their help and support in conducting the lab experiments. We are grateful to Johannes Abeler, Peter Andre, Anke Becker, Volker Benndorf, Jonathan Chapman, Felix Chopra, Julian Detemple, Thomas Dohmen, Armin Falk, Christina Felfe, Guido Friebe, Andreas Grunewald, Florian Hett, David Huffman, David Poensgen, Frederik Schwerter, Erik Snowberg, Uwe Sunde, Ferdinand von Siemens, and Matthias Wibral, as well as seminar participants at DICE, Frankfurt School, Goethe University Frankfurt, University of Konstanz, University of Lausanne, ESA Helsinki 2024, and the Behavioral Measurement Conference 2024 for helpful comments and suggestions. We gratefully acknowledge financial support from the Diligentia Stiftung and the Deutsche Forschungsgemeinschaft (550273651). 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 taking, patience, and the welfare of others, i.e., so-called “social” preferences. Variation in economic preferences has long been recognized as an important explanatory determinant for a wide array of differences in behaviors and real-life outcomes across many social, economic, and institutional domains. For example, risk preferences are associated with sorting into occupational choices 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), investment in risky assets (Dohmen et al., 2011), or even crime behavior (Epper et al., 2022). Time preferences are associated with health (Kirby and Petry, 2004; Chabris et al., 2008), education (Golsteyn et al., 2014; Backes-Gellner et al., 2021), savings (Sutter et al., 2013), credit card borrowing (Meier and Sprenger, 2010; Golsteyn et al., 2014), and wealth (Epper et al., 2020). 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), self-selection into public service (Friebel et al., 2019), and support for redistribution (Epper et al., 2024). Together, these results highlight the importance of accurate and precise, i.e., reliable measurement of economic preferences.

Traditionally, economists rely on the “revealed-preference approach” (Samuelson, 1938; Sen, 1973) by which preferences can be measured — are revealed — through individual choices made in controlled, incentivized decision-making situations. By varying the choice set, different types of preferences can be measured (e.g., ambiguity vs. risk, reciprocity vs. inequity aversion, etc.). While incentivized choice experiments — in the following, called *behavioral measures* — are a key methodological benchmark for measuring economic preferences, their complexity and resource intensity impose significant limitations. This becomes particularly pronounced when the aim is to measure preferences in large, representative populations. As a consequence, economists have started to move away from behavioral preference measures in recent years and replace them by survey instruments that are so-called “experimentally” or “behaviorally validated.” Key examples are the Global Preference Survey developed by Falk et al. (2018, 2023) measuring a range of economic preferences in the domains of risk, patience, as well as social preferences, Cavatorta and Schröder (2019) considering ambiguity attitudes, Fallucchi et al. (2020) and Buser et al. (2021) measuring competitiveness, Enke et al. (2022) for moral universalism, Schudy et al. (2024) for truth telling, and Epper and Mitrouchev (2025) for inequality aversion and altruism.

The perceived advantage of survey instruments is that they are highly portable and can be quickly deployed in different field settings, adapting better to specific contexts and circumstances

than incentivized choice experiments. However, surveys face their own challenges: responses may be prone to reference bias (Heckman and Kautz, 2021; Kautz et al., 2014; Lira et al., 2022), self-serving biases (Gino et al., 2016), and inattention and strategic motives can cause imprecise responses (Dohmen et al., 2011). Furthermore, survey questions are not incentive-compatible by design, which makes economists particularly skeptical about their reliability. By behaviorally validating the employed survey instruments, i.e., by selecting those items that have been shown to correlate significantly with corresponding behavioral preference measures, these concerns are considered to be alleviated, corroborating survey measures as a valid alternative measuring instrument.

The behavioral-validation approach raises a number of questions. Typically, the validation is based only on one sample of participants in one specific context (e.g., students in one country). Surveys, however, are then implemented on a global, often representative, scale. To what extent does the behavioral validation hold outside its original context? Are survey items as predictive for behavioral measures as in the original validation, or does this depend on cultural and country-specific contexts? Next, even if survey items are shown to be significantly correlated, they might not be the best items in terms of explanatory power and hence would not have been selected if the survey instrument had been validated in a different culture or context. Does the behavioral validation depend on where it is conducted? In other words, if, for example, Falk et al. (2023) had done their behavioral validation in China instead of Germany, would the Global Preference Survey look different? Finally, all of the above considerations rely at least implicitly on the assumption that behavioral preference measures are the superior instrument, sometimes also considered the gold standard. To what extent is this a valid assumption? Ultimately, how do we best measure economic preferences, and on what basis do we assess the performance of one instrument relative to another?

In this paper, we address the above questions focusing on the Global Preference Survey (GPS) and the corresponding Preference Survey Module (PSM) of Falk et al. (2018, 2023), which cover several key economic preferences and have been widely used in recent years. We replicate their behavioral validation on almost 2,000 student participants in five different countries — China, Colombia, Iran, Kenya, the U.S. — employing the exact same procedures as in the original validation study in Germany (Falk et al., 2023). That is, we employed the same experimental games and comprehensive battery of more than 100 survey questions. We consider students because the original validation did, allowing us to draw direct comparisons. Moreover, the variation in preferences among students across these six countries, including Germany, is shown not to be different from the variation observed in representative samples of the same countries in Falk et al. (2018).

Our results show that many GPS and PSM survey items perform well in predicting behavior in incentivized choice experiments; however, coefficients vary and are not always sizable. Across all preferences, quantitative survey items, i.e., hypothetical versions of the corresponding choice

experiment, consistently outperform qualitative items, i.e., self-assessments.¹ For negative reciprocity in most countries and several preferences in Kenya, performance is found to be relatively poorer compared to the other preferences.

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 quantitative items are consistently selected in all countries, the qualitative items vary between countries. As qualitative items contribute relatively little to explanatory power, however, the loss due to their variation is limited. Intriguingly, even though qualitative GPS items are outperformed by quantitative items, they are still selected among the best top-three single-item survey modules, based on their performance in all six countries. A very similar result is obtained when employing a Random Forest procedure.

Finally, Principal Component Analysis (PCA), both within and across all preferences, reveals that incentivized behavioral measures and qualitative survey measures — while related — “live in separate worlds.” They always load on distinct components, with quantitative items joining the incentivized behavioral measure. Intriguingly, cross-preference clustering (for social preferences, and also for risk and time) is observed only for behavioral and quantitative measures and almost never for qualitative ones. These results suggest that the two instruments capture rather different aspects of the same latent construct of an economic preference, casting doubt on the use of survey instruments as an easy substitute for behavioral measures. The results are further corroborated by comparing the predictive validity of behavioral vs. survey measures for both cross-country and cross-individual heterogeneity. While cross-country heterogeneity (such as Long-term Orientation, Human Development Index, or Corruption Index) seems to be better captured by behavioral measures, survey measures are found to be more predictive for differences between individuals (such as gender, smoking habits, or financial investments). To what extent these correlations reflect true relationships and hence suggest one measure to be superior to the other, unfortunately, our data does not allow us to answer. We believe, however, that it is important to show that performances in predicting behavior and outcomes differ.

Our results relate to a small number of studies that have started to replicate the behavioral validation of GPS measures in different contexts. 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 are driven by the fact that qualitative items may simply be noisier. Next, neither

¹For example, the quantitative item for measuring risk preferences is a multiple price list using the staircase method with hypothetical payments. The qualitative item is a self-assessment about the general willingness to take risks.

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. Our larger, multi-country study provides a more comprehensive picture: while there is important variation, by and large, the measures do better than expected — however, mostly due to the performance of quantitative survey items.

Chapman et al. (2025) is closest to our approach as they also use much larger samples to draw robust conclusions. Their focus is primarily on qualitative self-assessments. They document, in line with our results, that qualitative GPS items are only weakly correlated with incentivized elicitations. Further, qualitative items are sometimes more strongly correlated with the behavioral measure of another preference. We do not find this in our data. For all preferences, the qualitative GPS item is always most strongly correlated with its corresponding behavioral measure (cf. Table 3). Still, our PCA results corroborate the conclusion that qualitative items, while referring to the same latent construct, seem to capture different components than incentivized elicitations. A very similar message of both papers is therefore that behavioral validations, at least when relying only on qualitative self-assessments, may not be as “valid” as they are thought to be.

That quantitative and qualitative measures seem to capture different dimensions of an economic preference has also been shown by Frey et al. (2017) in the case of risk preferences. Our results suggest that this pattern is actually more general, as it can also be observed for patience and social preferences. Whether this implies that incentivized preference elicitations are the better measure, as Chapman et al. (2025) argue, or qualitative self-assessments, as Frey et al. (2017) do, is something we cannot resolve based on the currently available evidence. We believe that this requires more extensive investigations in the future. The goal of these analyses should be to help economists identify how (most likely, depending on the context and the particular research question) they can best measure economic preferences.

Altogether, our paper makes three important contributions to the literature. First, we document that many GPS and PSM survey items are 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 large samples drawn from multiple countries representing various continents and cultures. Second, this extensive dataset allows us to develop alternative survey modules that are informed by the observed variation in measures across cultural contexts. Third, employing a comprehensive set of survey questions and behavioral measures for multiple preferences enables us to closely examine both within- and across-construct associations between preferences assessed by these two approaches, thereby illuminating how the two methods capture distinct parts of individuals’ preference structures.

Our results highlight a number of critical gaps that warrant further investigation in future re-

search. We hope our research lays the groundwork for other researchers to join, e.g., by utilizing the study materials we develop in their own contexts across various countries, thus advancing the research agenda toward a comprehensive validation and development of truly global measures of economic preferences.

The paper is organized as follows. In the next section, we explain our main study design and measurement of economic preferences based on the GPS and PSM. Section 3 contains the results from the behavioral validations in the five different countries. Section 4 explores the deeper relationship between behavioral and survey measures, e.g., using PCA. Section 5 compares the predictive power of both measures for cross-country and cross-individual heterogeneity. Finally, Section 6 concludes.

2 Study Design and Measurement

Behavioral validations of survey questions have become increasingly popular in economics over the last years. We focus our analysis on the Global Preference Survey (GPS) developed by Falk et al. (2018, 2023). To ensure comprehensive coverage of our research questions, we implemented the complete study package described in Falk et al. (2023) following the exact same procedures as in their original study.

In order to conduct the experiments in different countries, we first screened active experimental economic laboratories across the globe and then selected the most compatible one in a given country, ensuring it met the standards of typical experimental economics laboratories and could comply with the particular aspects of our study. In particular, we required participants to have no prior experience in economic experiments, to come to the lab twice at the same time, one week apart, and, depending on their choices in the experiment, to be paid in one year’s time. In all cases, participants were recruited from the student body of a public, comprehensive university comparable to the University of Bonn, Germany, where Falk et al. (2023) had conducted their experiments. In China we conducted our experiments at the FEEL lab at Xiamen University; in Colombia at the REBEL lab at the University of Rosario, Bogota; in Iran at the Business Research Lab at the University of Tehran; in Kenya at the Busara Lab with students from the University of Nairobi; and in the United States at the EBEL lab at the University of California, Santa Barbara. In total, 1932 students (395 from China, 397 from Colombia, 389 from Iran, 383 from Kenya, 368 from the US) participated in our study, i.e., the sample size in each country is close to Falk et al. (2023).

Following the standard practice of translation and back-translation by native or bilingual speakers, we created the English master version based on the original German codes in zTree, and translated them into Mandarin, Farsi, and Spanish. As English is the official language in Kenya and the US, we conducted our study in both countries in English. All experiments and surveys were

programmed in oTree (Chen et al., 2016).

To minimize spillovers between incentivized behavioral measures and survey measures, we followed the procedures of Falk et al. (2023) and invited participants twice, with one session containing choice experiments on risk and time discounting preferences combined with survey items on social preferences, and another session containing choice experiments on social preferences combined with survey items on risk and time preferences. The two sessions were one week apart. Further, to mitigate potential order effects, sessions were organized in a way that in the first session, half of the sample participated in experiments on risk and time preferences, whereas the other half participated in experiments on social preferences.

We calibrated monetary incentives for each country using the 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). The local lab managers were actively engaged in this process to ensure comparable incentives across countries. In the original study conducted in Germany in the winter of 2010/2011, 100 experimental points were valued at 0.8 Euros. To account for inflation and cross-country differences in purchasing power, we calibrated the conversion rates in our study so that 100 experimental points corresponded to 4 Chinese yuan, 3,000 Colombian pesos, 150,000 Iranian rial, 30 Kenyan shillings, and 1 US dollar. The resulting average experimental points earned over both sessions are 8,728, 8,335, 8,142, 7,805, and 8,821 in China, Colombia, Iran, Kenya, and the US, respectively. The amounts correspond well in real terms to the average earnings in Falk et al. (2023).

For the time-discounting experiments in which participants have to choose between a payment today and a payment in 12 months, we utilized established payment methods in the local labs, in particular digital transfers via known platforms in the different countries. Participants were assured that they would receive their delayed payments with the institutional guarantee offered by the local lab. They were further informed that future payments will be inflation-adjusted.

On average, a session lasted about 100 minutes. Table B1 in Appendix B1 presents the implementation timeline and key demographic characteristics, such as age and gender, of participants across all countries. Figure B1 complements this information by showing the distribution of participants' field of study. As can be seen, participants in our study cover a diverse range of subjects.

2.1 Incentivized Choice Experiments

In the following, we briefly describe the incentivized choice experiments used for measuring the various preferences. More detailed information is provided in the Online-Appendix.² We retained the exact same games, instructions, order, and perfect-stranger-matching protocol in the games

²Original instructions are available at <https://nextcloud.flex.uni-frankfurt.de/index.php/s/ZHSRXyiFo3GcEfM>

involving strategic interaction as described in Falk et al. (2023). To reduce measurement errors, two variations of a choice 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 behavioral measure of a given preference is calculated by averaging the two choices in each version (see Table D1 in Appendix D).

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 points (i.e., the rows in which participants switched from preferring the lottery to the safe payment) in these two multiple price lists.³

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 points (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 either 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. Participants were given a list of well-established international and domestic charitable organizations in a given country. Participants could then designate a charity organization to which they wanted the money to be donated. We elicited altruism with the amount participants transferred to charity.⁴

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

³In case of multiple switching points, we use the average of the switching points as the final measure. We do the same for time preferences.

⁴Participants 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.

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 the 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.

As there were no control questions in the original experiments of Falk et al. (2023), we added control questions ourselves to ensure that participants understood the choice experiments and to be able to control for their level of comprehension in the analysis (cf. Online-Appendix). We always tried to ensure that all participants understood the instructions of each choice experiment before making their decisions. In particular, every participant was provided the correct answer to a 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. As explained 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. Falk et al. (2023) included 188 qualitative and quantitative survey items 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 by the authors.

We refined the survey battery to obtain a final set of 120 qualitative and quantitative survey items. 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 deleted these questions. In the Online-Appendix, we provide a detailed description of our iteration process and display all questions that remained as well as those that were excluded. All GPS and PSM items are, of course, still included in our study (see Table D2 in Appendix D).

To compare the performance of behavior- versus survey-based preference measures in explain-

ing 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 behavior and outcomes outside the lab. Finally, we also elicited demographics such as age, gender, cognitive ability, and socioeconomic status at the end of the study.

3 Behavioral Validation of the GPS and PSM in China, Colombia, Iran, Kenya, and the US

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 significantly predict behavior in the experiments in each country, as well as the combined data from all five replication countries.

Before doing so, we document that the cross-country variation in preferences in our student sample is comparable, and statistically indistinguishable, from the variation observed in the corresponding GPS sample. While we do not claim (nor expect) that students from a given country reveal exactly identical preferences to those from a representative sample, Figure B2 in the Appendix illustrates that the variation in preferences in these samples is indeed very similar. The Figure shows close alignment between the distributions of student and representative samples, indicating similar distributional properties, not only similar means but also comparable spread and shape, across the entire distribution of preferences. Non-parametric Leven’s tests reveal no significant differences in the cross-country variation between the two samples ($p > 0.2$) for any preference. The finding aligns with results in Snowberg and Yariv (2021) and corroborates the use of student samples to address our research questions.

For each preference, we estimate OLS regressions of the standardized behavioral measure on the corresponding GPS survey items, with all variables standardized within each country. Throughout the analysis, we focus on the sample of participants who answered the control questions correctly.⁵ Figure 1 shows the results by preference and country. The first coefficient at the top of each

⁵Results do not depend on this. Further, as there were no control questions in Falk et al. (2023), the sample from Germany includes all participants.

panel presents results from the original study in Germany, serving as a benchmark.⁶ It is calculated based on data from the replication package published by Falk et al. (2023). The subsequent coefficients correspond to China, Colombia, Iran, Kenya, and the US, respectively. The final coefficient in each panel shows results for all replication countries combined with country fixed effects, with all variables standardized across countries. The complete regression results are reported in Table C1 in Appendix C.

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 positive and 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, and in some cases, it is rather small. Results for the quantitative items on risk taking, patience, and altruism (R1, D1, and A11) show relatively strong and statistically significant coefficients across all countries. The coefficients of these items, particularly those related to risk taking and altruism, are remarkably similar across countries. The results align with prior evidence regarding the performance of quantitative survey items (Bauer et al., 2020; Kosfeld and Sharafi, 2023). In contrast, the qualitative survey items generally show more variation in their predictive power, with the trust question (T16) being an exception. However, the overall performance of these qualitative survey items appears considerably better than in the aforementioned studies.⁷

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 as well as combined.

For qualitative survey items, we observe notable differences across samples and preferences. The risk-related qualitative question does not meet the 95% CI criterion in any country except Colombia, while the time preference question does so in China, Iran, and the US. The trust question meets the criterion in all countries except Kenya, whereas the altruism question only meets it in China, and the positive reciprocity questions meet it in China and Colombia. In the Appendix, we provide 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 using the full sample without restricting the sample to students who answered the control questions correctly,

⁶GPS questions for negative reciprocity were later modified and therefore did not appear in the original validation study.

⁷For 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 at the five percent level.

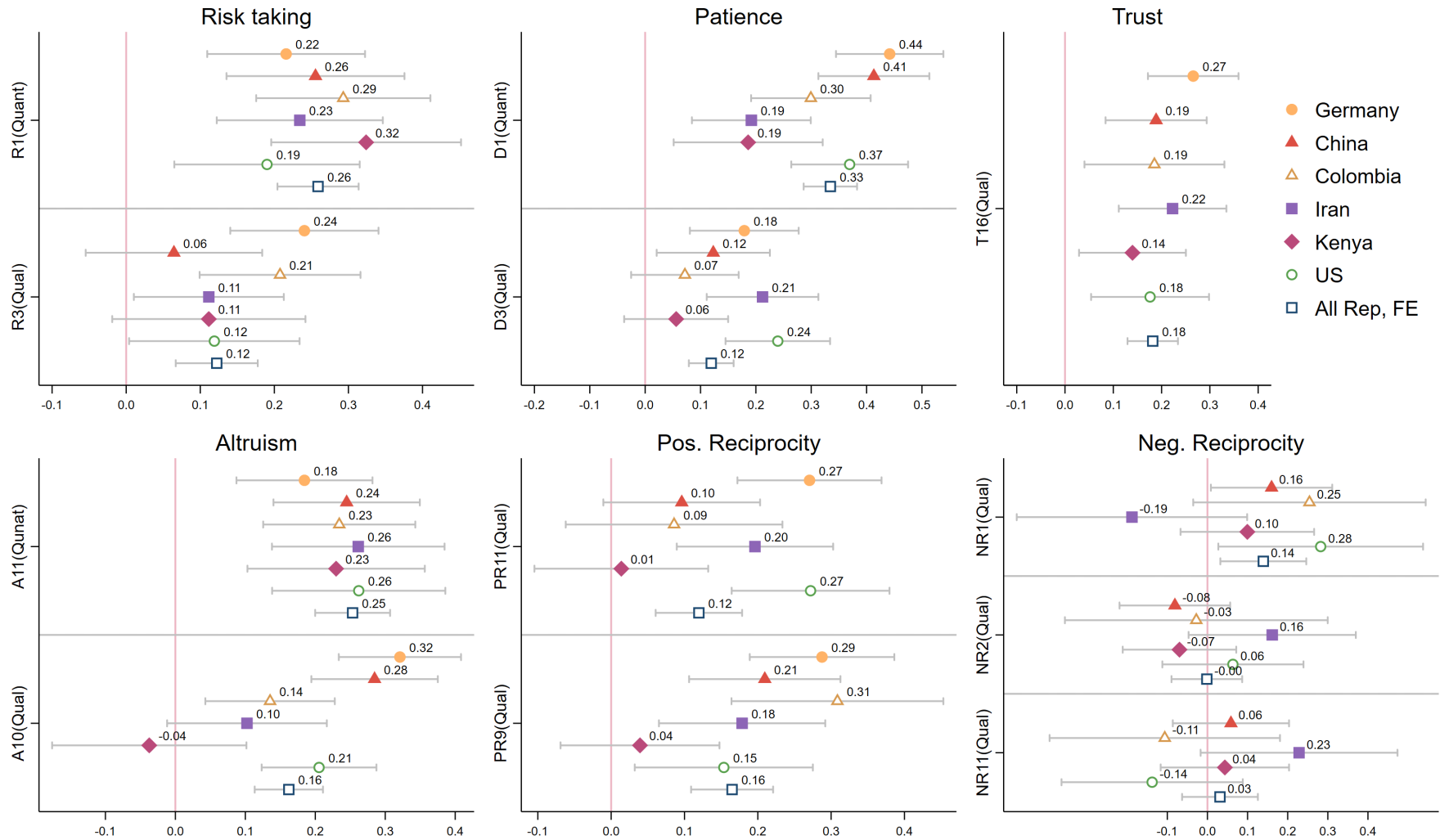


Figure 1: **Plot of regression coefficients of GPS items.** Coefficients from OLS regressions with 95% confidence intervals. The sample includes only participants who answered control questions correctly. See the Online-Appendix for similar results based on the full sample. For negative reciprocity, the 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. Regression results are presented in Table C1 in the Appendix.

and (cf. Online-Appendix) the results are very similar.

Comparing the size of coefficients between countries, we find that China, the US, and Germany appear to be very close in the domains of risk taking, patience, trust, altruism, and positive reciprocity. As for negative reciprocity, we are unable to draw inferences because the corresponding GPS items were modified after the original study.

Based on the GPS items, Falk et al. (2018) calculate the *GPS measure* for each preference, which is a weighted sum of both quantitative and qualitative survey items, with weights determined by the corresponding OLS regressions from the prior validation study (see Appendix D2). Following this methodology, we construct the GPS measure of each preference for every country and all replication countries combined using the original weights, and explore to what extent the GPS measures are associated with behavioral preference measures elicited from the experiments.

Figure 2 shows the results. As shown in the plot, the weighted GPS measures show consistent associations with the behavioral measures from the incentivized choice experiments, with the exception of negative reciprocity and an overall somewhat lower performance in Kenya. Moreover, the variation in coefficients across countries decreases, in particular for risk taking, trust, and altruism. Table C2 in the Appendix reports details of the corresponding OLS regressions results. We also check for robustness by considering the full sample of participants and show that results are again similar (cf. Online-Appendix).

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 A1 in the Appendix confirms this expectation, demonstrating that most PSM items are significantly associated with behavior in the lab (see Table C3 for regressions). Results based on the full sample are also similar (cf. Online-Appendix).

We summarize our first result as follows.

Result 1: *Most GPS and PSM items are significantly associated with behavior in incentivized choice experiments in the five replication countries, except for negative reciprocity in most countries. The performance of quantitative survey items is consistently better than the performance of qualitative survey items. There exist substantial differences in the magnitude of the observed associations between the replication countries and the original study in Germany. Overall, the lowest*

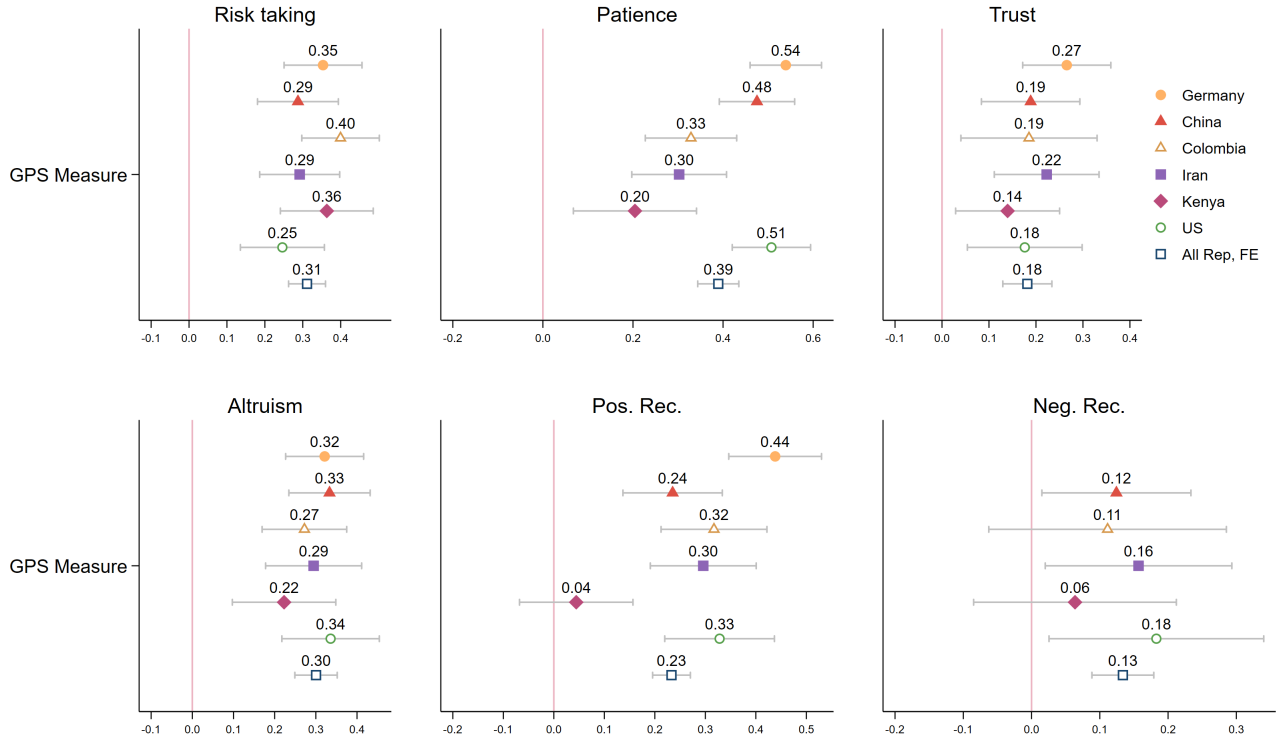


Figure 2: **Plot of regression coefficients of GPS measures.** Coefficients from OLS regressions with 95% confidence intervals. Sample includes only participants who answered control questions correctly. GPS measures are weighted measures combining both GPS items, with weights determined by the corresponding 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. Regression results are presented in Table C2 in the Appendix.

association for most preferences is found in Kenya.

3.3 Item Selection Based on Linear Combinations

Even if most GPS and PSM items are significantly correlated with the corresponding behavioral preference measure, this does not imply that they are necessarily the best items to explain behavior in incentivized choice experiments in a given country. Our next question is, therefore, 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 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 behavioral preference measure on all possible permutations of two survey items in the respective preference domain. Using

adjusted R^2 , this stage identifies the combinations of survey questions with the highest explanatory power for the choices made in the corresponding experiments. This technique is supposed to yield 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 combinations identified in different countries are identical to each other and to the one in Germany, and additionally, whether survey items identified with data pooled from all replication countries are identical to those in Germany.

Table 1 presents the results sorted by preferences. We document, for each preference, the selected combination of survey items in Germany (which serves as the benchmark), China, Colombia, Iran, Kenya, the US, all replication countries combined, and all six 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 candidates for explaining behavior in experiments, except for positive reciprocity in Colombia and negative reciprocity in both Colombia and Kenya. 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 survey item.⁸

This leads us to explore an important question: to what extent do the various qualitative survey items selected in different countries differ from each other in terms of their explanatory power for behavior in experiments? In other words, how much predictive power do we lose by switching from one qualitative item to another? To answer this question, we pool the data from all countries except Germany and plot, for each preference, the sorted adjusted R^2 of all possible combinations of two survey items. As can be seen in Figure 3, 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

⁸What would be the outcome, if we included *all* 120 survey items (independent of their target preference) in the item selection procedure for a given behavioral measure? To assess the robustness of the results, we reran the procedure across all available survey items for all preference domains. Results show that the overall pattern remains largely unchanged, indicating stability in the identified questions for each preference and the strongest association of survey items to their corresponding behavioral measure. Only two minor deviations emerged: in the case of trust, the model selected a risk-related qualitative item instead of the original trust survey question; in the case of positive reciprocity, an altruism-related qualitative item was selected. In both cases, however, the substitution yielded only a marginal improvement (smaller than 0.003) in model fit. Results are available from the authors upon request.

Preference	Items	Countries	Item Number
Risk taking	1. Multiple price lists	DE, CN, IR, KE, US, CO All_Rep, All	R2
	2. How likley: I like taking risks	All_Rep, All	R15
	3. Willingness to take risks in traffic	IR, US	R10
	4. Willingness to take risks	DE	R3
	5. Willingness to take risks compared with others	CO	R4
	6. Imagine you win 100.000 Euro in a lottery. You can participate in the lottery using the whole amount you won or only a part of it	CN	R16
Number of observations: DE: 382(360); CN: 385; IR: 341; KE: 258, US: 344, CO: 315, All (replication combined): 1643, All: 2025 (2003)			
Patience	1. Multiple price lists	DE, CN, US, CO All_Rep, All	D2
	2. Willingness to wait	IR, All_Rep, All	D3
	3. Staircase measure	IR, KE	D1
	4. Willingness to wait in comparison with others?	DE	D4
	5. Willingness to wait wrt. financial investments	US	D6
	6. I am a person who often does not keep their own good resolutions	CN	D18
	7. How much money do you save per month?	KE	D19
Number of observations: DE: 382(360); CN: 391; IR: 370; KE: 360, US: 368, CO: 371, All (replication combined): 1860, All: 2242 (2220)			
Trust	1. Hypothetical first-mover behavior	DE, CN, IR, KE, US, CO All_Rep, All	T24
	2. Willingness to trust people in your professional environment	KE, All_Rep, All	T6
	3. Willingness to trust strangers	CO	T7
	4. Willingness to trust people in your neighborhood	CN	T8
	5. People have only the best intentions	DE	T16
	6. Nowadays one cannot rely on anyone anymore.	US	T18
	7. Most people would take advantage of you when they have the chance	IR	T20
Number of observations: DE: 382(360); CN: 352; IR: 323; KE: 317, US: 268, CO: 218, All (replication combined): 1478, All: 1860 (1838)			
Altruism	1. Hypothetical donation game	DE, CN, IR, KE, US, CO, All_Rep, All	A11
	2. I am willing to donate time and money to charity, even if I do not profit from that directly	CN, IR, US, CO, All_Rep, All	A13
	3. Share when it comes to good causes	DE	A10
	4. I am a person who would give their shirt off their back to help others	KE	A18
Number of observations: DE: 382(360); CN: 378; IR: 351; KE: 274, US: 333, CO: 310, All (replication combined): 1646, All: 2028 (2006)			
Pos. Reciprocity	1. Hypothetical second-mover behavior	DE, CN, IR, KE, US, All_Rep, All	PR16
	2. Willingness to return a favor when it comes to strangers	CN, CO, All_Rep; All	PR7
	3. Thank-you gift in the lost in the city scenario	DE, IR, CO	PR9
	4. When someone does me a favor, I am willing to return it	US	PR11
Number of observations: DE: 382(360); CN: 352; IR: 323; KE: 317, US: 268, CO: 218, All (replication combined): 1478, All: 1860 (1838)			
Neg. Reciprocity	1. Minimum acceptable offer in ultimatum game	DE, CN, IR, US, All_Rep, All	NR10
	2. Willingness to punish compared with others	IR, CO	NR3
	3. If someone harms me on purpose, I will try to give that person a taste of their own medicine (NR14)	CN, All_Rep	NR14
	4. Willing to punish unfair behavior even if this is costly	DE	NR1/2
	5. Willingness to punish someone who treats you unfairly	CO	NR1
	6. Willingness to punish people in your hometown	KE	NR5
	7. Willingness to punish unfair behavior when it comes to your neighborhood	KE	NR9
	8. When someone puts me into a difficult situation, I will do the same to them	All	NR12
	9. When someone treats me in a bad way, I do not just let it go	US	NR18
Number of observations: DE: 382(360); CN: 321; IR: 186; KE: 151, US: 200, CO: 109, All (replication combined): 967, All: 1349 (1327)			

Table 1: Best survey items across countries. This table shows the best combination of survey items based on the iterative linear item selection procedure in Falk et al. (2023). Sample includes only participants who answered control questions correctly. Blue-colored items are GPS items.

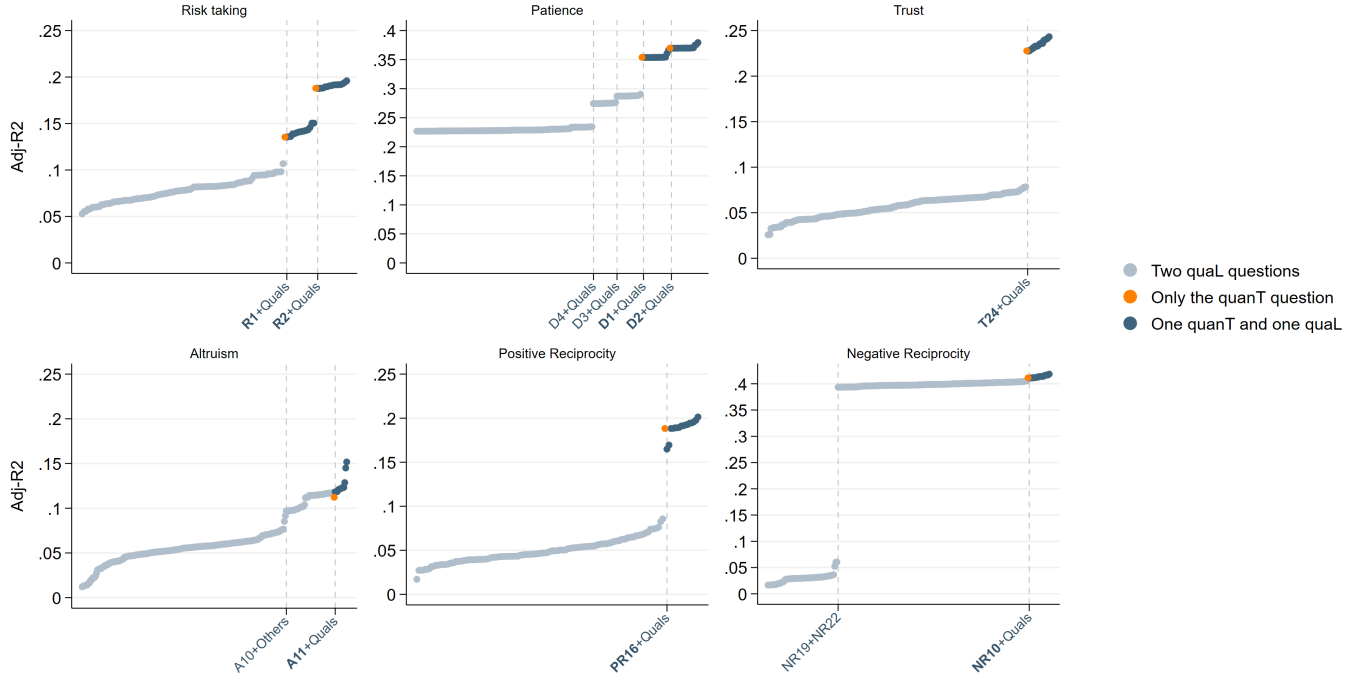


Figure 3: **Sorted adjusted R^2 .** This figure presents the adjusted R^2 s from OLS regressions of the incentivized measure on different combinations of two-items in the pooled dataset. The bold-faced items represent quantitative questions.

combination for the incentivized behavior. While the iterative linear item selection procedure selects the combination of two items that has the highest predictive power, it ignores whether or not the marginal increase in adjusted R^2 between different models is meaningful. 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 second result is as follows.

Result 2: *Deploying the same item selection procedure as in the original validation study, we find that the quantitative items are consistently selected in each country, whereas the qualitative items vary between countries. However, the overall contribution of qualitative items to explaining behavior in the experiments is relatively minor. This suggests that once the quantitative question is included, the choice of the accompanying qualitative question has a limited impact.*

3.4 Best Survey Items for Approximating Behavior in All Countries

Our rich cross-country data allow us to address the question of how to best approximate behavioral preference measures by survey items, also on a different, more global scale. Suppose a researcher wants to employ surveys to measure economic preferences in all six of our countries. Can we identify survey items that predict corresponding lab behavior well in each of these countries? Do

these items include the GPS items? In this section, we provide an answer to these questions.

For each preference, we present single-item and multiple-item modules derived from two types of analysis. Firstly, we run univariate OLS regressions of the standardized behavioral measure on each corresponding standardized survey 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 the behavioral measure at least at a five-percent level *in all six countries*. If no item meets this criterion, we shortlist items that are significantly associated with behavior in five out of the six countries (with the exception of negative reciprocity, in which no item meets the five out of six criterion). In this way, we aim to 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 next 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.⁹

Table 2 shows the results. Given the strong performance of quantitative items, they are always included in the selected questions for all preferences. However, we also identify qualitative items. For negative reciprocity, only two questions pass our inclusion criteria. Notably, when examining single items, the GPS questions (printed in blue) — both the quantitative and qualitative items — are retained in the shortlist for all preferences, with the exception of negative reciprocity. This indicates that the GPS items are generally good candidates for measuring preferences across all countries, 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, the GPS items are at least effective within the scope of those put together by Falk et al. (2023) and included in our analysis.

Secondly, we develop a multiple-item module derived from the application of the Random Forest procedure. Feature selection involves identifying the most informative explanatory variables (i.e., a subset of relevant features) to describe a response variable (Mullainathan and Spiess, 2017). This method helps reduce the influence of redundant or irrelevant features, as well as those strongly correlated in the data, while preserving predictive power. For risk and time preferences, we identify two quantitative items, the multiple price list and the staircase measure. As the correlation between these two items exceeds 0.7 in our dataset, we exclude the longer measure and retain only the staircase measure before applying the Random Forest procedure to the pool of questions for these preferences, using variable importance scores to guide the selection of the most relevant items.

Results of the Random Forest procedure are also shown in Table 2. As can be seen, GPS items are again included for all preferences with the exception of negative reciprocity. To evaluate the

⁹In the pooled dataset, we again standardize all experimental measures and corresponding survey questions across countries.

Preference	Module	Item	Countries	Adj R^2	MSE test
Risk taking	Single-item	Quant - R2: Multiple price lists	CN; CO; DE; IR; KE; US	0.188	0.859
		Quant - R1: Switching row from staircase measure	CN; CO; DE; IR; KE; US	0.135	0.924
		Qual - R3: Willingness to take risks	CN; CO; DE; IR; KE; US	0.095	1.03
	Multiple-item (Random Forest)	R1: Switching row from staircase measure R15: How well do the following statements describe you as a person? I like taking risks. R16: You win 100K EUR in a lottery. How much of it would you invest in the risky yet profitable lottery?		0.153	0.905
Patience	Single-item	Quant - D2: Multiple price lists (D2)	CN; CO; DE; IR; KE; US	0.37	0.713
		Quant - D1: Switching row from staircase measure	CN; CO; DE; IR; KE; US	0.354	0.759
		Qual - D3: Willingness to wait	CN; CO; DE; IR; US	0.287	0.991
	Multiple-item (Random Forest)	D1: Switching row from staircase procedure D3: Willingness to wait D19: How much money do you save per month?		0.366	0.736
Trust	Single-item	Quant - T24: First-mover behavior in trust game	CN; CO; DE; IR; KE; US	0.227	0.869
		Qual - T16: I assume that people have only the best intentions.	CN; CO; DE; IR; KE; US	0.064	0.986
		Qual - T17: In general, one can trust other people.	CN; CO; DE; IR; KE; US	0.064	0.985
	Multiple-item (Random Forest)	T16: I assume that people have only the best intentions. T17: In general, one can trust other people. T24: 1st mover decision in hypothetical trust game		0.246	0.890
Altruism	Single-item	Quant - A11: Hypothetical donation game	CN; CO; DE; IR; KE; US	0.112	0.920
		Qual - A13: I am willing to donate time and money to charity, even if I do not profit from that directly.	CN; CO; DE; IR; US	0.114	0.907
		Qual - A10: Willingness to give to good causes	CN; CO; DE; IR; US	0.097	0.937
	Multiple-item (Random Forest)	A10: Willingness to give to good causes A11: Hypothetical donation game A13: I am willing to donate time and money to charity, even if I do not profit from that directly.		0.157	1.10
Pos. Reciprocity	Single-item	Quant - PR16: 2nd mover behavior in hypothetical trust game	CN; CO; DE; IR; KE; US	0.188	0.951
		Qual - PR9: Thank-you gift in the lost in the city scenario	CN; CO; DE; IR; US	0.064	1.03
		Qual - PR11: When someone does me a favor	CN; CO; DE; IR; US	0.057	1.00
	Multiple-item (Random Forest)	PR9: Thank-you gift in the lost in the city scenario PR11: When someone does me a favor PR16: 2nd mover behavior in hypothetical trust game		0.202	0.824
Neg. Reciprocity	Single-item	NR12: When someone puts me into a difficult situation, I will do the same to them.	CN; DE; IR; US	0.401	1.08
		NR14: If someone harms me on purpose I will try to give that person a taste of their own medicine.	CN; DE; IR; US	0.402	1.10
	Multiple-item (Random Forest)	NR19: I absolutely dislike being the fool. NR21: Hypothetical scenario: cutting wages NR22: It is important to me to be respected by others.		0.393	0.766

Table 2: Best survey items in all countries. This table reports alternative (single-item and multiple-item) survey modules for each preference. Sample includes only participants who answered control questions correctly. 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 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.

relative performance of all 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 (70%) and test (30%) 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. In the single-item module, the quantitative items (except for altruism) 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.

We suggest researchers make a situation-specific decision when it comes to which items to use, balancing predictive power and generalizability with parsimony.

4 Exploring the Deeper Relationship between Behavioral and Survey Measures of Economic Preferences

So far, our analysis has taken the incentivized behavioral preference measures as a benchmark. To further explore the relationship between behavioral measures on the one hand and survey measures on the other hand, we analyze the within- and across-preference correlations between both types of measures and conduct principal component analyses of all measures together in this section. We thus no longer treat the incentivized behavioral measures as *explanandum* but consider all (incentivized, quantitative, qualitative) measures as ex-ante equally relevant variables.

4.1 Within- and Across-Preference Correlations Between Incentivized Behavioral Measures and Qualitative GPS Questions

Table 3 shows the correlation coefficients between the incentivized behavioral preference measures and each of the GPS qualitative survey items using the pooled data from all countries. Two observations emerge. Firstly, for all six preferences, the GPS qualitative item is most strongly correlated with its corresponding incentivized behavioral measure, as indicated by the darker blue shading along the diagonal. Secondly, these qualitative items also show meaningful correlations with behavioral measures from related preference domains. This is particularly evident among social preferences: trust, altruism, and positive reciprocity are all positively correlated with each other, while negative reciprocity is negatively correlated with trust, altruism, and positive reciprocity.

Qualitative GPS Question	Incentivized Behavioral Measure					
	<i>Risk</i>	<i>Patience</i>	<i>Trust</i>	<i>Altruism</i>	<i>Pos. Reciprocity</i>	<i>Neg. Reciprocity</i>
Risk taking	0.228*** (0.024)	0.022 (0.02)	0.136*** (0.026)	0.058** (0.027)	0.017 (0.027)	0.047** (0.023)
Patience	0.043* (0.026)	0.280*** (0.02)	0.070*** (0.025)	0.052** (0.026)	0.076*** (0.027)	-0.024 (0.023)
Trust	0.040* (0.024)	-0.007 (0.019)	0.201*** (0.023)	0.156*** (0.024)	0.137*** (0.023)	-0.010 (0.023)
Altruism	0.007 (0.021)	0.013 (0.02)	0.064** (0.025)	0.284*** (0.023)	0.163*** (0.024)	-0.051** (0.026)
Pos. Reciprocity	0.002 (0.025)	-0.017 (0.019)	0.180*** (0.025)	0.146*** (0.026)	0.231*** (0.025)	-0.068** (0.027)
Neg. Reciprocity	0.027 (0.028)	-0.026 (0.022)	-0.071** (0.03)	-0.113*** (0.028)	-0.126*** (0.03)	0.146*** (0.034)

Table 3: **Correlation coefficients of incentivized behavioral elicitations and (GPS) qualitative self-assessments.** ***, **, * denote statistical significance at the 1%, 5%, and 10% level.

Chapman et al. (2025) document much weaker correlations between behavioral and qualitative GPS measures of a given preference. One possible reason could be that participant samples differ — in our case, students from six countries across the globe, in their case, a representative sample of US citizens. Additional studies are clearly needed to provide more information.

4.2 Principal Component Analysis of All Measures

Next, to uncover the deeper, structural relationship between preference measures, we conduct a principal component analysis (PCA) of *all* measures together. We first do this for each preference separately. Subsequently, we pool all the data to analyze the cross-preference structure of our measures. PCA is a widely used linear dimensionality reduction technique that transforms original variables into a new set of principal 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.

Table 4 contains the PCA results for each preference separately based on the data from all six countries, restricted to participants who answered control questions correctly.¹⁰ Results for the full sample are reported in the Online-Appendix. In line with established practice in the literature, we document 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,

¹⁰As before, we include all participants for Germany.

Preference	Principal components	Corresponding survey questions	Fraction of Expl. Var.
A: Risk taking	Willingness to take risks	R3-R8, R12, R15-R16	0.298
	<i>Quantitative measures for risk</i>	Exp. incen. measure, R1, R2	0.123
	Risk in financial investments	R13-R16	0.097
	Risk in daily activities	R9-R11	0.094
B: Patience	Willingness to wait in financial situations	D5-D6, D9-D11, D14-D15	0.135
	Willingness to wait	D3-D5, D7-D8, D17	0.126
	<i>Quantitative measures for patience</i>	Exp. incen. measure, D1, D2	0.116
	Not keeping resolutions	D11, D14, D16-D18	0.107
	Compensation for giving up a vacation trip	D7-D8, D12-D13	0.056
C: Trust	Willingness to trust	T1-T8, T13-T14, T17	0.255
	Most people are fair	T20, T21	0.086
	Building trust in other people and their intentions	T10-T13, T16	0.076
	Most people act in their own best interest	T22, T23	0.071
	<i>Quantitative measures for trust</i>	Exp. incen. measure, T24	0.060
	Willingness to rely on strangers	T15, T18-T19	0.049
	Willingness to lend to strangers	T5, T7, T9, T10	0.048
D: Altruism	Willingness to share with people	A1-A10	0.276
	Willingness to do good if I expect it returns	A12, A15-A17, A20	0.104
	<i>Quantitative measures for altruism</i>	Exp. incen. measure, A10-A11, A13-A14, A17	0.084
	Be an unselfish person	A18, A19, A21	0.078
	Hours of volunteering	A17, A22-A23	0.062
E: Pos. reciprocity	Willingness to return a favor	PR1-PR8	0.292
	I help someone who has helped me before	PR11-PR15	0.183
	Thank-you gift to return a favor	PR9-PR10, PR16	0.099
	<i>Quantitative measures for positive reciprocity</i>	Exp. incen. measure, PR14, PR16	0.081
F: Neg. reciprocity	Willingness to punish people with unfair behavior	NR2-NR9	0.235
	Willingness to punish people who treated you unfairly	NR1, NR11-NR14, NR16, NR18	0.210
	Dislike being taken for a fool and pushed around	NR15, NR17, NR19-NR20, NR22	0.104
	<i>Quantitative measures for negative reciprocity</i>	Experimental incen. measure, NR10	0.052
	Fairness in hypothetical scenario	NR17, NR21	0.047

Table 4: **Principal components of 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.

we apply a varimax rotation to the resulting components (Furr, 2021). On average, we identify between four and five components per preference with trust as the exception of having seven components. 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.

As can be seen in Table 4, the component in bold print primarily aggregates clusters of qualitative self-assessments related to an individual’s willingness to either take risks, patience, trust others, share with others, return a favor, or punish those who behave unfairly, with time preference as an exception of having willingness to wait in specific conditions such as financial decisions. Next, a distinct component (in italics in the table) bundles the quantitative measures of each preference, i.e., the incentivized behavioral measure and its hypothetical survey versions. These quantitative measures are always clustered together into one component for each preference. As the identified components are orthogonal to each other, this suggests that quantitative and qualitative preference measures contribute distinct information to the data structure, with each type of measure capturing different aspects of the latent construct, i.e., preference.

The results documented in Table 4 are remarkably consistent with evidence from Frey et al. (2017), who document a very similar separation of quantitative and qualitative measures in the case of risk preferences. Our results suggest that such separation may be more general and applies similarly to patience as well as social preferences. The results are also reminiscent of — and in a certain sense corroborate — the asymmetric performance of quantitative and qualitative survey items in predicting behavior in the incentivized preference measures, as documented in the previous section.

To uncover the cross-preference structure of our measures, we next conduct a PCA of the complete dataset, i.e., of all measures of all preferences pooled together. Because restricting the sample to participants who answered all control questions correctly would reduce the sample size too much for this type of analysis, we work with the full sample. For interpretability and visualization, we restrict the extraction to components with eigenvalues greater than two, which yields a total of ten principal components. This criterion produces a compact yet informative representation of the underlying covariance structure of the various preference measures. Figure 4 shows the results. On the vertical axis, we list all measures with loadings strictly greater than 0.2 on any of the ten different components. On the horizontal axis, we list the ten principal components ranging from components that capture risk taking (on the left) to negative reciprocity (on the right).¹¹

Three important observations can be made in Figure 4. Firstly, quantitative measures of a given preference, incentivized or hypothetical, always load on the same component. This finding, once again, corroborates observations made above. Secondly, for the quantitative measures of different

¹¹Note that components are not ordered by the fraction of explained variance in this case.

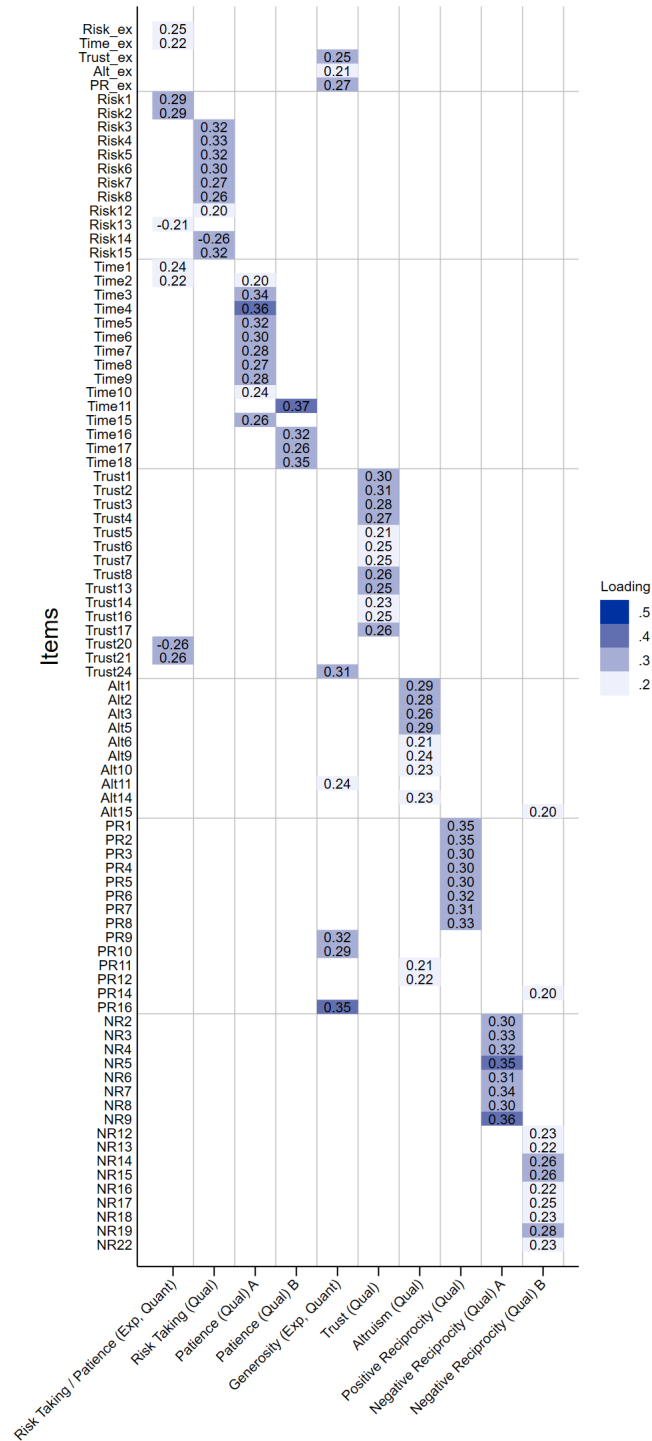


Figure 4: **Principal Component Analysis on all items** The figure shows the varimax rotated factor loadings for 10 components extracted from incentivized behavioral measures and all survey questions measuring risk taking, patience, trust, altruism, positive reciprocity, and negative reciprocity. Only loadings with absolute values ≥ 0.2 are displayed. The legend presents the middle number of the range, i.e., 0.2 represents loading between 0.15 and 0.25.

preferences, we see important clustering across preferences. For example, risk taking and patience go together, loading on the same component. Here, we also see loadings of two trust items, which is intuitive as trust is a particular form of risk taking behavior. Next, the quantitative measures of trust, altruism, and positive reciprocity also cluster together. This finding is very similar to Chapman et al. (2023), who document the same kind of clustering of social preferences based on incentivized behavioral measures. Following their notation, we label the common component *generosity*. Intriguingly and thirdly, we almost never see any of these cross-preference clustering for the qualitative preference measures. All qualitative measures of a given preference always load on one or at most two distinct components and very rarely go with their quantitative measures.¹² In other words, if a quantitative measure of a given preference has two options to go with — its corresponding qualitative measures or a quantitative measure of another preference — it typically goes with the latter. The fact that cross-preference clustering is observed for quantitative and not for qualitative preference measures appears striking and, to the best of our knowledge, has not been documented before. It, once again, suggests that quantitative and qualitative preference measures — while related — seem to “live in separate worlds.”

We summarize our results as follows.

Result 3: *Qualitative GPS items are most strongly correlated with the corresponding incentivized behavioral measure of a given preference. PCA analysis suggests that quantitative measures (incentivized or hypothetical) and qualitative measures capture distinct dimensions of the same latent construct of an economic preference. Cross-preferences clustering is only observed for quantitative preference measures (risk-time, generosity) and almost never for qualitative measures.*

5 Explaining Heterogeneity between Countries and between Individuals

Economic preferences have long been recognized as important explanatory factors for a wide array of differences in behaviors and real-world outcomes across social, economic, and institutional domains. In this section, we revisit this approach by leveraging our dataset to systematically compare the extent to which different preference measures are able to explain variation in relevant individual- and country-level behaviors and outcomes.

One important caveat is, of course, that our data comprise only six countries and individual-level variation in student samples is naturally limited. Furthermore, the analysis relies on the assumption that a deeper (causal) relationship between preferences on the one hand and behaviors

¹²The two only exceptions are the qualitative component of altruism, where we see two items of positive reciprocity, and a qualitative component of negative reciprocity with loadings on two items of altruism and of positive reciprocity.

and outcomes on the other hand truly exists. Unfortunately, it is often unclear whether this assumption is valid. Notwithstanding, we believe that the results from the following analysis are useful, both for exploring the connection between preferences and outcomes and for comparing different methods to measure preferences. We first consider country-level differences; subsequently, we analyze differences at the individual level.

5.1 Heterogeneity between Countries

At a macro level, prior research has demonstrated that aggregated economic preferences, when measured across national populations, can account for significant cross-country variation in key economic, institutional, and social outcomes. Building on this, we extend our analysis by integrating a set of existing country-level indices that are both theoretically and empirically associated with core preference dimensions such as risk tolerance, patience, trust, and altruism (Algan and Cahuc, 2013; Falk et al., 2018; La Porta et al., 1996). More precisely, we examine how these macro indices relate to behavior- and survey-based measures of economic preferences, thereby linking our individual-level preference measures from the different countries to broader cultural and institutional contexts. Specifically, our analysis incorporates long-term orientation and uncertainty avoidance from Hofstede (2001), the Human Development Index (HDI) compiled by the United Nations Development Program, the Corruption Perceptions Index (CPI) from Transparency International¹³, and the Environmental Performance Index (EPI) developed by Block et al. (2024).

Figure 5 documents the empirical relationship between the different macro indices and our aggregated preference measures. For each index and measure, we transform country-level z-scores on a scale from zero to 100 and plot all six countries in a diagram with preference measures on the X-axis and macro indices on the Y-axis. A perfect match between the indices and a preference measure would result in the observations lying on the 45-degree line of the corresponding graph. The left panel of Figure 5 shows results for the incentivized behavioral measure of a given preference, the right panel for the corresponding qualitative GPS item.

Several notable patterns can be identified from this analysis. With respect to uncertainty avoidance (uppermost panel), e.g., neither of the two measures of risk preferences captures the heterogeneity between countries very well. With respect to the other indices and preferences, results look better, in particular for the incentivized behavioral measure. For example, for long-term orientation (second panel from the top), we see that the behavioral measure of patience is better able to capture country differences in this index than the qualitative GPS item. The same holds for the relationship between patience and HDI (third panel from the top), where the behavioral measure suggests that countries that score high in HDI tend to exhibit greater patience. For the qualitative

¹³The CPI measures perceived levels of public sector corruption — that is, how corrupt people think a country’s government and institutions are. The index scores countries on a scale from 0 (very corrupt) to 100 (very clean).

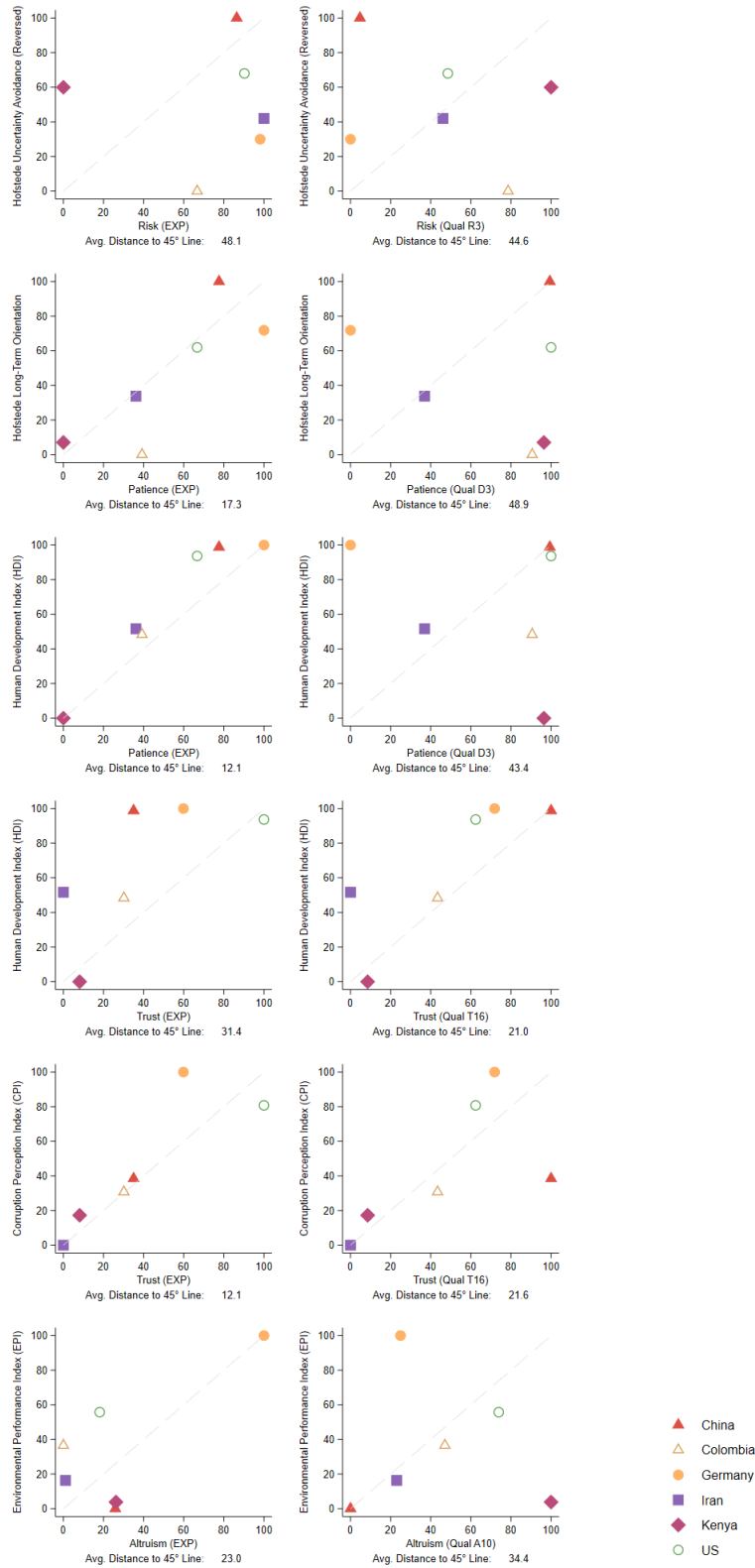


Figure 5: **Preference measures and country-level differences.** The figure shows the relationship between aggregated preference measures (X-axis) and various country-level indices (Y-axis). Left panel: incentivized behavioral measure (EXP), right panel: qualitative GPS question (SUR).

measure, this relationship is not found. With respect to HDI and trust (third panel from the bottom) both preference measures do rather similarly, while for trust and CPI (second panel from the bottom) the behavioral measure again does somewhat better. Finally, with respect to environmental performance (EPI, undermost panel), neither measure produces a clear picture. According to the behavioral measure, all countries except Germany score relatively low on altruism, while they exhibit somewhat larger variation in EPI. For the qualitative measure, countries appear well aligned with the notable exception of Germany and Kenya.

In sum, the results suggest that incentivized behavioral preference measures seem to align more closely with country-level differences in selected macro indices.

5.2 Heterogeneity between Individuals

We next turn to individual-level differences. Prior literature consistently documents that women tend to be more risk averse and more altruistic than men (Byrnes et al., 1999; Croson and Gneezy, 2009). Do we find similar gender differences in our data, and do differences depend on how preferences are measured?

The upper two panels of Figure 6 show OLS regression coefficients of risk preferences and altruism on gender, based on the pooled data from all replication countries. Results confirm a significant gender difference in risk taking for both the incentivized behavioral measure and the qualitative GPS item, while a significant gender difference in altruism is found only for the qualitative survey item.

In our study, we also elicited self-reported information about participants' risky behaviors in the domains of health (smoking) and finance (investing in securities). We further collected data about saving behavior, social capital (number of friends), participation in social organizations and pro-environmental behavior. See the Online-Appendix for details on the questions asked. The lower panels of Figure 6 show coefficients from OLS regression of these individual real-life behaviors on various preference measures (cf Table C4 for regression results), again based on the pooled replication data. In particular, we consider risk preferences for smoking and investment (Anderson and Mellor, 2008; Dohmen et al., 2011), patience for saving (Sutter et al., 2013), trust for social capital (Dohmen et al., 2009), and altruism for participation in social organizations as well as pro-environmental behavior (Cárdenas et al., 2013; Lades et al., 2021).

Two observations stand out. Firstly, survey-based preference measures are consistently and significantly associated with real-life behaviors across the various domains. Secondly, incentivized behavioral preference measures are either uncorrelated or only weakly associated with real-life behaviors, with the exception of patience and savings. These findings are consistent with evidence from several other studies (e.g. Coppola, 2014; Lönnqvist et al., 2015; Verschoor et al., 2016; Frey

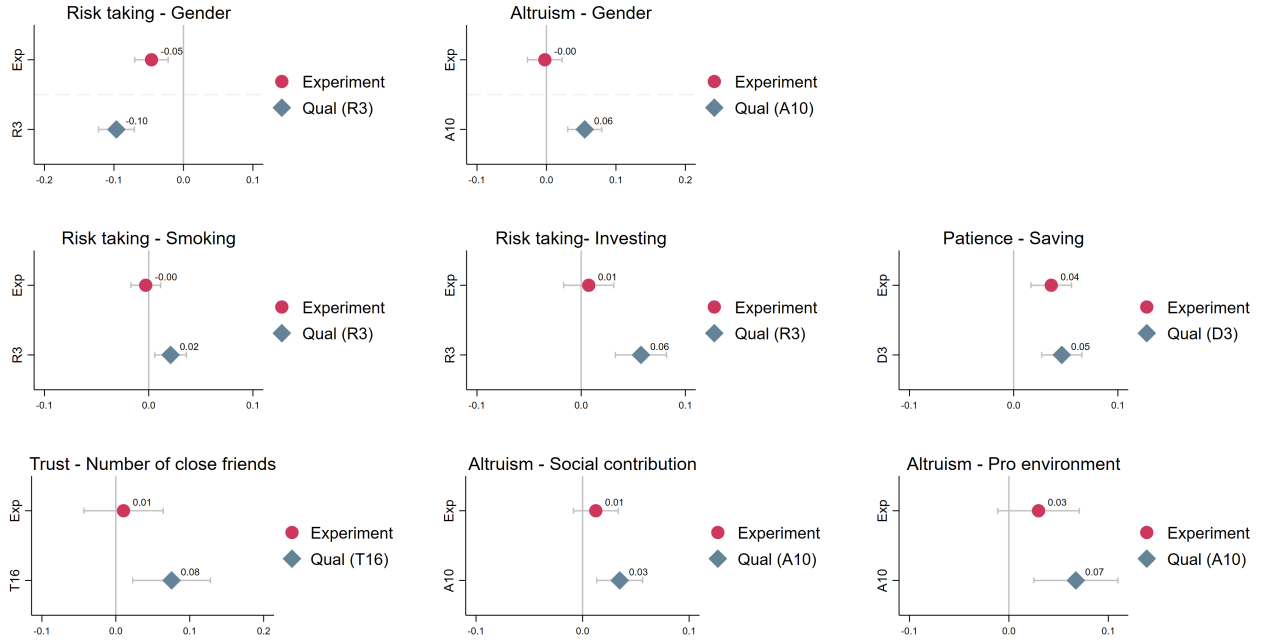


Figure 6: Preference measures and individual-level differences. In the upper panel of this plot, we document OLS regression coefficients of gender on various measures of risk taking and altruism. In the lower part, including six plots, we report OLS regressions of various preference measures on real-life behavior. The data is a pooled sample of China, Iran, Kenya, Colombia, and the US, focusing only on participants who answered control questions correctly. We control for country fixed effects. Regression results are presented in Table C4 in the Appendix.

et al., 2017; Pinger, 2017; Galizzi and Navarro-Martinez, 2019; Charness et al., 2020; Rafäi et al., 2023; Chapman et al., 2025) suggesting that qualitative survey items tend to outperform behavioral measures in predicting individual-level differences in real-life behavior. We consider these findings important but emphasize that the extent to which they allow us to draw conclusions regarding the *predictive validity* of different preference measures remains an open question.¹⁴ On the one hand, it may be that qualitative self-assessments truly capture the latent construct of a given preference as a determinant of a particular form of real-life behavior in a better way. On the other hand, it is possible that the significant correlation between qualitative self-assessments and real-life behavior — in particular in the case of self-reported real-life behavior — is a consequence of other factors (e.g., omitted variable bias). To provide an example, the fact that self-reported pro-environmental behavior correlates with self-assessed altruism but not with incentivized dictator-game giving, yields two equally plausible, yet opposing conclusions: either altruism drives pro-environmental behavior and the measure based on self-assessment better captures this; or altruism does *not* drive pro-

¹⁴The same holds, of course, for conclusions about the observed differences in explaining country-level heterogeneity.

environmental behavior and the measure based on incentivized behavior better captures this, while the observed correlation is an outcome of, e.g., social-desirability or self-image concerns. Obviously, we need more research to provide a final answer to this question.

We summarize our last finding as follows:

Result 4: *Incentivized behavioral preference measures appear to align more closely with aggregated, country-level differences, while survey-based preference measures seem to be better predictors for individual-level differences in (self-reported) real-life behaviors.*

6 Conclusion

With this paper, we aim to contribute to making progress in our understanding of how to best measure economic preferences. Traditionally, economists rely on incentivized behavioral elicitation via carefully designed choice experiments following a revealed-preference approach. Other social science disciplines, instead, such as psychology and sociology, employ carefully designed survey instruments typically based on self-assessments. The pros and cons of both measurement approaches lie at hand. Choice experiments provide an instrument that is almost ideally close to the construct of an economic preference. What else could it mean other than that someone *prefers* X over Y when they *choose* X instead of Y when both options are available? Yet, choice experiments are costly, often requiring substantial effort and financial resources to implement, particularly when being conducted in the field. Their instructions can be difficult for participants to fully understand, and it remains unclear to what extent the choices made under the specific framings and laboratory environments designed by researchers can be generalized to real choices and decisions, which are often not made in such a deliberate and reflective manner. Survey instruments, on the other hand, are highly portable tools that can be quickly deployed in the field and are often better adapted to specific contexts and circumstances. Still, they face their own challenges, such as reference bias or self-serving bias. In addition, they are not incentive compatible.

By “experimentally” or “behaviorally validating” survey instruments, recent papers in economics have tried to combine the best of both methods. The resulting modules are easily deployed in the field, and they are thought to capture what they are supposed to do by documenting significant correlations with behavior in incentivized choice experiments. Our results in this paper show that this approach may not be as straightforward as it seems.

Based on the Global Preference Survey (GPS) and the corresponding Preference Survey Module (PSM) of Falk et al. (2018, 2023), which we replicate with about 2,000 student participants in China, Colombia, Iran, Kenya, and the US, we show that many of the GPS and PSM items are significantly correlated with behavior in incentivized choice experiments in the five different countries. Among the best top-three single-item survey modules that predict behavior in all countries

are always the two GPS questions, both the quantitative and the qualitative items. The only exception is negative reciprocity, where we do not identify any of the GPS questions. However, the size of the correlations varies across countries, and for some preferences, it is not very large. Moreover, quantitative survey items, i.e., hypothetical versions of the experiment, typically perform much better than qualitative items relying on self-assessments.

The difference between quantitative and qualitative items becomes even more pronounced when we explore the underlying structure of survey items and behavioral measures both within and across preferences. To investigate this underlying structure, we draw on our unique dataset comprising incentivized behavioral measures, quantitative survey items, and extensive qualitative self-assessment data. Quantitative items and their corresponding incentivized behavioral measure always load together on the same principal component, while qualitative items are always separated. Across preferences the same holds; here, we further find that quantitative and behavioral items cluster even across preferences (risk taking together with time discounting, generosity combining altruism, trust, and positive reciprocity), which is never seen for qualitative items.

Thus, while survey instruments based on qualitative self-assessments *are significantly related* to incentivized behavioral measures, they *are not the same* in the sense of capturing the same dimensions of their underlying latent construct. This casts doubt on the use of one instrument as an easy substitute for the other. A very similar result has been shown by Frey et al. (2017) for risk preferences. Our results suggest that this finding holds more generally.

We hope our results motivate economists, and social scientists more generally, to investigate more extensively the relationships between different methods for measuring economic preferences. Every science needs good measurement. Our study raises a number of questions we still need to answer in order to make sufficient progress in this respect: In what conditions is one measure indeed superior to another, and on what factors does this depend? What are the *methodological* benefits and costs of each instrument, rather than their advantages in terms of implementation? Can we identify (and define ex-ante) questions and contexts for which we would need behavioral measures, and others where self-assessments are the appropriate instrument? What is the best behavioral measure for a given economic preference? Currently, the economic literature is far away from good answers or anything close to a consensus on these problems.

References

- ABDI, H. AND L. J. WILLIAMS (2010): “Principal component analysis,” *WIREs Computational Statistics*, 2, 433–459.
- ALGAN, Y. AND P. CAHUC (2013): “Trust and growth,” *Annu. Rev. Econ.*, 5, 521–549.
- 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.
- BLOCK, S., J. W. EMERSON, D. C. ESTY, A. D. SHERBININ, AND Z. A. WENDLING (2024): “2024 Environmental Performance Index: Technical appendix,” .
- 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.
- BUSER, T., M. NIEDERLE, AND H. OOSTERBEEK (2021): “Can competitiveness predict education and labor market outcomes? Evidence from incentivized choice and survey measures,” *Review of Economics and Statistics*, forthcoming.
- 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.
- CAVATORTA, E. AND D. SCHRÖDER (2019): “Measuring ambiguity preferences: A new ambiguity preference survey module,” *Journal of Risk and Uncertainty*, 58, 71–100.
- 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., M. DEAN, P. ORTOLEVA, E. SNOWBERG, AND C. CAMERER (2023): “Econographics,” *Journal of Political Economy Microeconomics*, 1, 115–161.
- CHAPMAN, J., P. ORTOLEVA, E. SNOWBERG, L. YARIV, AND C. CAMERER (2025): “Reassessing qualitative self-assessments and experimental validation,” Tech. rep., National Bureau of Economic Research, 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.
- 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.
- ENKE, B., R. RODRIGUEZ-PADILLA, AND F. ZIMMERMANN (2022): “Moral universalism: measurement and economic relevance,” *Management Science*, 68, 3590–3603.
- EPPER, T., E. FEHR, H. FEHR-DUDA, C. T. KREINER, D. D. LASSEN, S. LETH-PETERSEN, AND G. N. RASMUSSEN (2020): “Time discounting and wealth inequality,” *American Economic Review*, 110, 1177–1205.
- EPPER, T., E. FEHR, K. B. HVIDBERG, C. T. KREINER, S. LETH-PETERSEN, AND G. NYTOFT RASMUSSEN (2022): “Preferences predict who commits crime among young men,” *Proceedings of the National Academy of Sciences*, 119, e2112645119.

- EPPER, T. F., E. FEHR, C. T. KREINER, S. LETH-PETERSEN, I. S. OLUFSEN, AND P. E. SKOV (2024): “Inequality aversion predicts support for public and private redistribution,” *Proceedings of the National Academy of Sciences*, 121, e2401445121.
- EPPER, T. F. AND I. MITROUCHEV (2025): “Measuring Hearts and Minds: A Validated Survey Module on Inequality Aversion and Altruism,” *Journal of Economic Inequality*, 23, 777–808.
- 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.
- FALLUCCHI, F., D. NOSENZO, AND E. REUBEN (2020): “Measuring preferences for competition with experimentally-validated survey questions,” *Journal of Economic Behavior & Organization*, 178, 402–423.
- FREY, R., A. PEDRONI, R. MATA, J. RIESKAMP, AND R. HERTWIG (2017): “Risk preference shares the psychometric structure of major psychological traits,” *Science Advances*, 3, e1701381.
- 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.
- GALIZZI, M. M. AND D. NAVARRO-MARTINEZ (2019): “On the external validity of social preference games: a systematic lab-field study,” *Management Science*, 65, 976–1002.
- 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 lifetime 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.
- HECKMAN, JAMES J., J. T. AND T. KAUTZ (2021): *Some Contributions of Economics to the Study of Personality*, New York, NY: Guilford Press.
- HOFSTEDE, G. (2001): *Culture’s consequences: Comparing values, behaviors, institutions and organizations across nations*, Sage publications.

- 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.
- 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.
- LA PORTA, R., F. LOPEZ-DE SILANES, A. SHLEIFER, AND R. W. VISHNY (1996): “Trust in large organizations,” .
- 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’BRIEN, 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.
- LÖNNQVIST, J.-E., M. VERKASALO, G. WALKOWITZ, AND P. C. WICHARDT (2015): “Measuring individual risk attitudes in the lab: Task or ask? An empirical comparison,” *Journal of Economic Behavior & Organization*, 119, 254–266.
- MEIER, S. AND C. SPRENGER (2010): “Present-biased preferences and credit card borrowing,” *American Economic Journal: Applied Economics*, 2(1), 193–210.
- MULLAINATHAN, S. AND J. SPIESS (2017): “Machine learning: an applied econometric approach,” *Journal of Economic Perspectives*, 31, 87–106.
- 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.

- RAFAÏ, I., T. BLAYAC, D. DUBOIS, S. DUCHÊNE, P. NGUYEN-VAN, B. VENTELOU, AND M. WILLINGER (2023): “Stated preferences outperform elicited preferences for predicting reported compliance with Covid-19 prophylactic measures,” *Journal of Behavioral and Experimental Economics*, 107, 102089.
- 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.
- SCHUDY, S., S. GRUNDMANN, AND L. SPANTIG (2024): “Individual preferences for truth-telling,” *CESifo Working Paper Series*, 11521.
- SEN, A. (1973): “Behaviour and the concept of preference,” *Economica*, 40, 241–259.
- SNOWBERG, E. AND L. YARIV (2021): “Testing the waters: Behavior across participant pools,” *American Economic Review*, 111, 687–719.
- 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 Replication Results on PSM

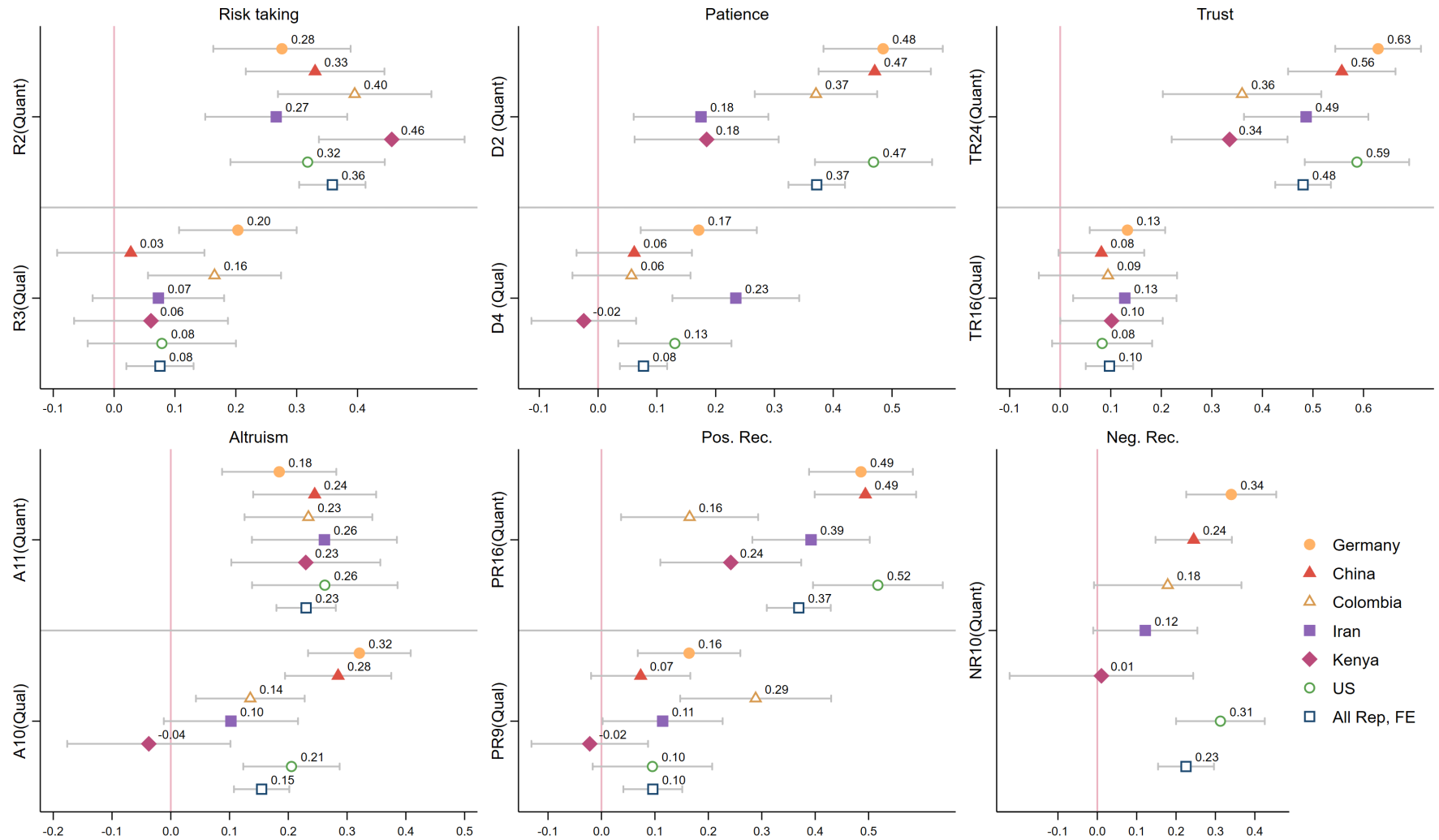


Figure A1: Plot of regression coefficients of PSM items. Coefficients from OLS regressions with 95% confidence intervals. The sample includes only participants who answered control questions correctly. For negative reciprocity we only include one quantitative item because the original one qualitative question on willingness to punish people has been split into two questions for the GPS. 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. Regression results are presented in Table C3 in the Appendix.

B Summary Statistics

B1 Study Participants

	China	Iran	Kenya	US	Colombia	All
Date	Oct 2023	April/May 2023	Nov/Dec 2023	April/May 2025	May 2025	
Place of study	Xiamen	Tehran	Nairobi	Santa Barbara	Bogota	
Female (%)	63.8	53.7	42.6	67.90	50.10	55.50
Age (Mean (S.D.))	19.9 (1.6)	22.7 (3.9)	20.9 (1.9)	20.0(2.4)	21.1 (5.1)	20.9 (3.4)
Bachelor	0.92	0.58	0.98	0.97	0.88	0.87
Master or above	0.08	0.41	0.18	0.03	0.12	0.13
Observations	395	390	383	368	397	1932

Table B1: **Demographic Information of Study Participants.** This table shows the time of replication and descriptive statistics for each country as well as for all countries together.

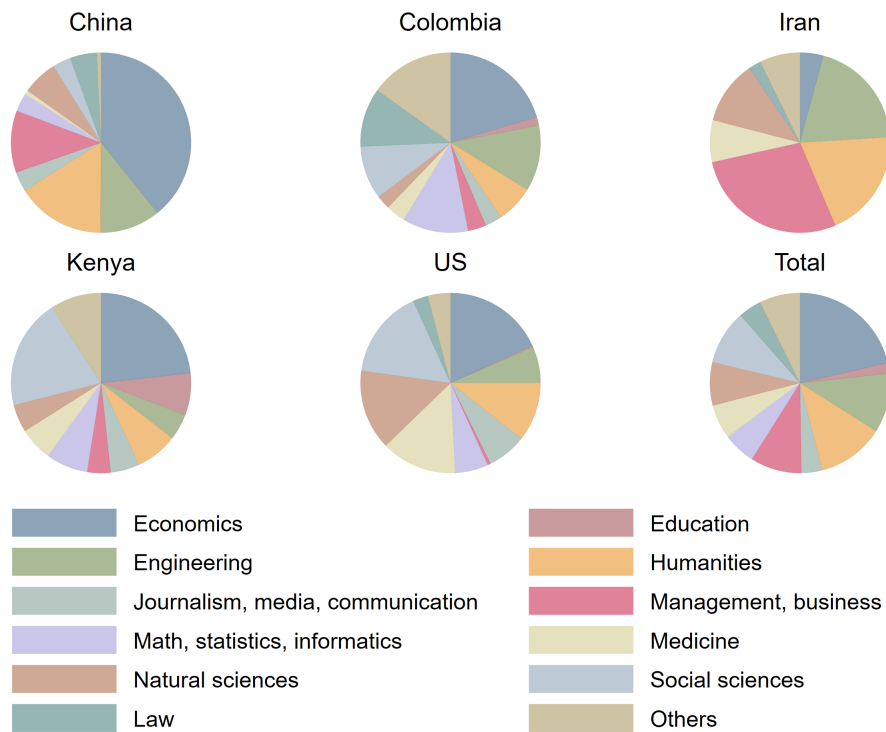


Figure B1: **Distribution of study participants' subject fields.** This pie chart presents the subject fields in each country as well as in the overall sample.

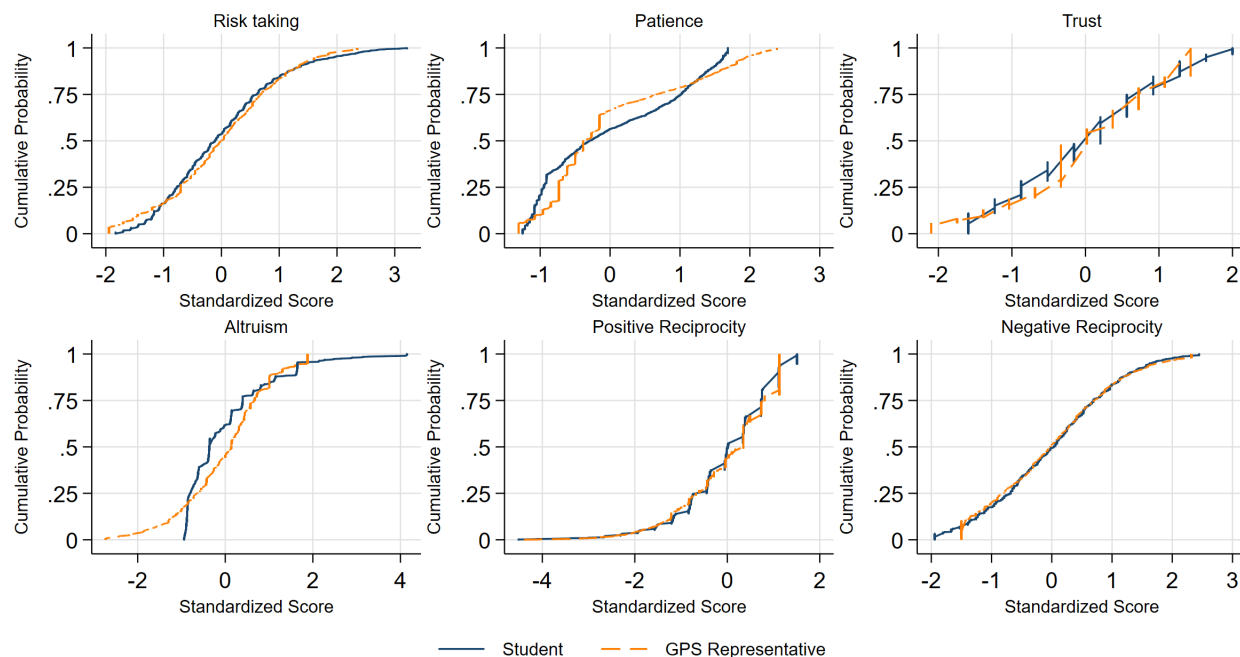


Figure B2: Distribution Comparison of Preferences Across Samples Cumulative distribution functions comparing student and GPS representative samples across preferences. Both samples include individuals from China, Colombia, Germany, Iran, Kenya, and the United States, and are standardized within these six countries. Non-parametric Leven's tests reveal no significant differences in the cross-country variation between the two samples ($p > 0.2$) for any preference.

C Regression Tables

C1 GPS Items, GPS Measure, and PSM Items

Preference	Items	Germany	China	Iran	Kenya	US	Colombia	Rep. combined FE
Risk taking	Staircase measure (R1)	0.216*** (0.054)	0.256*** (0.061)	0.234*** (0.057)	0.324*** (0.065)	0.190*** (0.064)	0.293*** (0.060)	0.259*** (0.028)
	Willingness to take risks (R3)	0.241*** (0.051)	0.064 (0.061)	0.112** (0.051)	0.112* (0.066)	0.119** (0.058)	0.208*** (0.055)	0.122*** (0.028)
			[-0.055,0.184]	[0.010,0.213]	[-0.019,0.242]	[0.004,0.234]	[0.099,0.316]	[0.067,0.178]
	Adjusted R-squared	0.140	0.079	0.079	0.113	0.057	0.168	0.148
	Number of observations	382	385	341	258	344	315	1643
Patience	Staircase measure (D1)	0.442*** (0.049)	0.413*** (0.051)	0.192*** (0.055)	0.186*** (0.068)	0.369*** (0.054)	0.299*** (0.055)	0.335*** (0.025)
	Willingness to wait (D3)	0.179*** (0.050)	0.123** (0.052)	0.212*** (0.051)	0.056 (0.048)	0.240*** (0.048)	0.072 (0.049)	0.119*** (0.021)
			[0.021,0.225]	[0.111,0.313]	[-0.038,0.150]	[0.145,0.334]	[-0.025,0.169]	[0.079,0.160]
	Adjusted R-squared	0.299	0.229	0.115	0.039	0.284	0.106	0.303
	Number of observations	382	391	370	360	368	371	1860
Trust	People have only the best intentions. (T16)	0.266*** (0.048)	0.189*** (0.053)	0.223*** (0.057)	0.140** (0.056)	0.176*** (0.062)	0.185** (0.074)	0.182*** (0.027)
			[0.084,0.293]	[0.111,0.334]	[0.029,0.250]	[0.054,0.298]	[0.040,0.330]	[0.129,0.234]
	Adjusted R-squared	0.070	0.032	0.042	0.018	0.026	0.030	0.060
	Number of observations	382	352	323	317	268	218	1478
Altruism	Hypothetical donation game (A11)	0.185*** (0.049)	0.245*** (0.053)	0.261*** (0.063)	0.230*** (0.064)	0.262*** (0.063)	0.234*** (0.055)	0.253*** (0.027)
	Share when it comes to good causes (A10)	0.321*** (0.044)	0.285*** (0.046)	0.102* (0.058)	-0.037 (0.071)	0.205*** (0.042)	0.135*** (0.047)	0.162*** (0.025)
			[0.194,0.375]	[-0.012,0.217]	[-0.176,0.102]	[0.123,0.287]	[0.043,0.228]	[0.113,0.211]
	Adjusted R-squared	0.175	0.183	0.091	0.055	0.133	0.091	0.110
	Number of observations	382	378	351	274	333	310	1646
Pos. Reciprocity	When someone does me a favor (PR11)	0.271*** (0.050)	0.096* (0.054)	0.196*** (0.054)	0.014 (0.060)	0.272*** (0.055)	0.086 (0.075)	0.120*** (0.030)
	Thank-you gift in the lost in the city scenario (PR9)	0.288*** (0.050)	0.210*** (0.052)	0.179*** (0.058)	0.039 (0.055)	0.154** (0.062)	0.309*** (0.073)	0.165*** (0.029)
			[0.106,0.313]	[0.065,0.292]	[-0.069,0.148]	[0.032,0.275]	[0.164,0.453]	[0.109,0.221]
	Adjusted R-squared	0.194	0.058	0.073	-0.004	0.122	0.117	0.070
	Number of observations	360	352	323	317	268	218	1478
Neg. Reciprocity	Punish someone who treats you unfairly (NR1)	N.A.	0.160** (0.077)	-0.188 (0.145)	0.099 (0.084)	0.282** (0.129)	0.254* (0.146)	0.139** (0.055)
			[0.009,0.311]	[-0.475,0.099]	[-0.067,0.266]	[0.027,0.537]	[-0.035,0.544]	[0.032,0.246]
	Punish someone who treats others unfairly (NR2)	N.A.	-0.081 (0.070)	0.161 (0.106)	-0.070 (0.072)	0.063 (0.089)	-0.028 (0.165)	-0.002 (0.045)
			[-0.219,0.057]	[-0.047,0.370]	[-0.211,0.072]	[-0.112,0.239]	[-0.355,0.300]	[-0.090,0.087]
	Take revenge at the first occasion (NR11)	N.A.	0.059 (0.074)	0.228* (0.124)	0.043 (0.081)	-0.138 (0.115)	-0.106 (0.145)	0.031 (0.048)
			[-0.086,0.203]	[-0.017,0.474]	[-0.117,0.203]	[-0.364,0.088]	[-0.393,0.181]	[-0.063,0.126]
	Adjusted R-squared	N.A.	0.020	0.029	-0.004	0.039	0.009	0.029
	Number of observations	382 (360)	321	186	151	200	109	967

Table C1: **Regression coefficients of GPS Items** Coefficients from OLS regressions with robust standard errors in parentheses and 95% confidence intervals in square brackets, based on participants who passed control questions.. Blue colored items are GPS items. Germany is the reference column. Black with boldface highlights coefficients that are statistically indistinguishable from those in Germany. Plain black indicates coefficients that are statistically different from those 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	US	Colombia	Rep. combined w FE
Risk taking	Staircase measure (R1) Willingness to take risks (R3)	0.095*** (0.014) [0.067,0.122]	0.083*** (0.016) [0.052,0.114]	0.077*** (0.014) [0.049,0.105]	0.081*** (0.014) [0.054,0.108]	0.066*** (0.015) [0.037,0.096]	0.095*** (0.012) [0.071,0.119]	0.311*** (0.025) [0.263,0.360]
	Adjusted R-squared	0.124	0.081	0.081	0.117	0.058	0.162	0.148
	Number of observations	382	385	341	258	344	315	1643
Patience	Staircase measure (D1) Willingness to wait (D3)	0.069*** (0.005) [0.059,0.079]	0.062*** (0.006) [0.051,0.073]	0.038*** (0.007) [0.025,0.052]	0.028*** (0.010) [0.009,0.047]	0.059*** (0.005) [0.049,0.069]	0.044*** (0.007) [0.030,0.057]	0.389*** (0.023) [0.344,0.435]
	Adjusted R-squared	0.289	0.225	0.092	0.040	0.255	0.107	0.298
	Number of observations	382	391	370	360	368	371	1860
Trust	People have only the best intentions. (T16)	0.266*** (0.048) [0.172,0.359]	0.189*** (0.053) [0.084,0.293]	0.223*** (0.057) [0.111,0.334]	0.140** (0.056) [0.029,0.250]	0.176*** (0.062) [0.054,0.298]	0.185** (0.074) [0.040,0.330]	0.182*** (0.027) [0.129,0.234]
	Adjusted R-squared	0.070	0.032	0.042	0.018	0.026	0.030	0.060
	Number of observations	382	352	323	317	268	218	1478
Altruism	Hypothetical donation game (A11) Share when it comes to good causes (A10)	0.006*** (0.001) [0.004,0.008]	0.001*** (0.000) [0.001,0.001]	0.000*** (0.000) [0.000,0.000]	0.000*** (0.000) [0.000,0.000]	0.005*** (0.001) [0.003,0.006]	0.000*** (0.000) [0.000,0.000]	0.301*** (0.026) [0.249,0.352]
	Adjusted R-squared	0.091	0.111	0.084	0.057	0.101	0.074	0.089
	Number of observations	382	378	351	274	333	310	1646
Pos. Reciprocity	When someone does me a favor (PR11) Thank-you gift in the lost in the city scenario scenario (PR9)	0.361*** (0.038) [0.286,0.437]	0.214*** (0.046) [0.125,0.304]	0.215*** (0.039) [0.139,0.292]	0.028 (0.036) [-0.043,0.099]	0.266*** (0.045) [0.178,0.354]	0.261*** (0.044) [0.175,0.348]	0.221*** (0.028) [0.166,0.276]
	Adjusted R-squared	0.195	0.053	0.076	-0.001	0.113	0.107	0.070
	Number of observations	360	352	323	317	268	218	1478
Neg. Reciprocity	Punish someone who treats you unfairly (NR1) Punish someone who treats others unfairly (NR2) Take revenge at the first occasion (NR11)	NA	0.065** (0.029) [0.008,0.122]	0.071** (0.031) [0.009,0.133]	0.027 (0.032) [-0.036,0.091]	0.088** (0.038) [0.012,0.164]	0.048 (0.038) [-0.027,0.124]	0.221*** (0.028) [0.166,0.276]
	Adjusted R-squared		0.012	0.019	-0.000	0.027	0.004	0.028
	Number of observations		321	186	151	200	109	967

Table C2: **Regression coefficients of GPS Measures.** Coefficients from OLS regressions with robust standard errors in parentheses and 95% confidence intervals in square brackets, based on participants who passed control questions. 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	US	Colombia	Rep. combined FE
Risk	Multiple price lists (R2)	0.276*** -0.057	0.330*** -0.058 [0.216,0.444]	0.266*** -0.059 [0.150,0.383]	0.456*** -0.061 [0.337,0.576]	0.318*** -0.064 [0.191,0.445]	0.395*** -0.064 [0.269,0.521]	0.359*** -0.028 [0.304,0.413]
	Willingness to take risk (R3)	0.203*** -0.049	0.027 -0.062 [-0.094,0.148]	0.073 -0.055 [-0.035,0.181]	0.061 -0.064 [-0.066,0.187]	0.078 -0.062 [-0.043,0.200]	0.165*** -0.056 [0.056,0.274]	0.075*** -0.028 [0.020,0.131]
	Adjusted R-squared	0.162	0.112	0.09	0.194	0.116	0.227	0.195
	Number of observations	382	385	341	258	344	315	1643
Time	Multiple price lists (D2)	0.485*** -0.052	0.471*** -0.049 [0.375,0.566]	0.175*** -0.058 [0.061,0.290]	0.185*** -0.062 [0.062,0.307]	0.469*** -0.051 [0.369,0.568]	0.371*** -0.053 [0.267,0.475]	0.372*** -0.025 [0.324,0.420]
	Willingness to wait in comparison with others? (D4)	0.171*** -0.05	0.061 -0.05 [-0.037,0.160]	0.234*** -0.055 [0.126,0.342]	-0.024 -0.045 [-0.113,0.065]	0.131*** -0.049 [0.034,0.227]	0.057 -0.051 [-0.044,0.157]	0.077*** -0.021 [0.037,0.117]
	Adjusted R-squared	0.34	0.25	0.112	0.029	0.289	0.149	0.31
	Number of observations	382	391	370	360	368	371	1860
Trust	First-mover behavior in trust game (T24)	0.629*** -0.043	0.557*** -0.054 [0.451,0.663]	0.487*** -0.063 [0.364,0.610]	0.335*** -0.058 [0.221,0.450]	0.587*** -0.052 [0.484,0.690]	0.360*** -0.08 [0.203,0.517]	0.481*** -0.028 [0.425,0.536]
	People have only the best intentions. (T16)	0.133*** -0.038	0.081* -0.043 [-0.003,0.166]	0.128** -0.052 [0.026,0.230]	0.102** -0.052 [0.000,0.203]	0.083* -0.05 [-0.016,0.182]	0.094 -0.069 [-0.042,0.231]	0.097*** -0.024 [0.051,0.144]
	Adjusted R-squared	0.452	0.329	0.267	0.136	0.344	0.132	0.265
	Number of observations	382	352	323	317	268	218	1478
Altruism	Hypothetical donation game (A11)	0.185*** -0.049	0.245*** -0.053 [0.140,0.350]	0.261*** -0.063 [0.138,0.385]	0.230*** -0.064 [0.103,0.357]	0.262*** -0.063 [0.138,0.386]	0.234*** -0.055 [0.126,0.343]	0.253*** -0.027 [0.200,0.307]
	Share when it comes to good causes. (A10)	0.321*** -0.044	0.285*** -0.046 [0.194,0.375]	0.102* -0.058 [-0.012,0.217]	-0.037 -0.071 [-0.176,0.102]	0.205*** -0.042 [0.123,0.287]	0.135*** -0.047 [0.043,0.228]	0.162*** -0.025 [0.113,0.211]
	Adjusted R-squared	0.175	0.183	0.091	0.055	0.133	0.091	0.11
	Number of observations	382	378	351	274	333	310	1646
Pos. Reciprocity	Second-mover behavior in trust game (PR16)	0.486*** -0.049	0.494*** -0.048 [0.399,0.589]	0.392*** -0.056 [0.282,0.502]	0.242*** -0.067 [0.110,0.374]	0.517*** -0.062 [0.396,0.639]	0.165** -0.065 [0.037,0.293]	0.369*** -0.031 [0.309,0.429]
	Thank-you gift in the lost in the city scenario (PR9)	0.164*** -0.049	0.073 -0.047 [-0.019,0.166]	0.114** -0.057 [0.002,0.227]	-0.022 -0.055 [-0.131,0.087]	0.095* -0.057 [-0.016,0.207]	0.289*** -0.072 [0.147,0.430]	0.096*** -0.028 [0.041,0.151]
	Adjusted R-squared	0.329	0.264	0.18	0.053	0.283	0.14	0.172
	Number of observations	360	352	323	317	268	218	1478
Neg. Reciprocity	Minimum acceptable offer (NR10)	0.340*** -0.058	0.245*** -0.049 [0.148,0.342]	0.122* -0.067 [-0.010,0.254]	0.01 -0.118 [-0.223,0.244]	0.313*** -0.057 [0.200,0.425]	0.179* -0.094 [-0.008,0.366]	0.225*** -0.036 [0.154,0.296]
	Adjusted R-squared	0.115	0.057	0.011	-0.007	0.109	0.026	0.049
	Number of observations	360	321	186	151	200	109	967

Table C3: Regression coefficients of PSM. Coefficients from OLS regressions with robust standard errors in parentheses and 95% confidence intervals in square brackets, based on participants who passed control questions. 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.

C2 Individual Real-life Behavior

Risk taking (N=1643)		Smoking	Investing in Securities	Gender
Experiment		-0.003 (0.007)	0.007 (0.012)	-0.046*** (0.012)
Adjusted R-squared		0.057	0.020	0.047
R1: Staircase measure		0.023*** (0.008)	0.022* (0.012)	-0.038*** (0.013)
Adjusted R-squared		0.062	0.022	0.045
R3: Willingness to take risks		0.021*** (0.008)	0.057*** (0.013)	-0.096*** (0.013)
Adjusted R-squared		0.060	0.032	0.069
R14: I avoid risky things		-0.039*** (0.008)		0.061*** (0.013)
Adjusted R-squared		0.071		0.053
R6: Willingness to take risks in financial investment			0.099*** (0.013)	-0.110*** (0.014)
Adjusted R-squared			0.056	0.075
Patience (N=1860)		Save money		
Experiment		0.036*** (0.010)		
Adjusted R-squared		0.041		
D1: Staircase measure		0.034*** (0.010)		
Adjusted R-squared		0.040		
D3: Willingness to wait		0.046*** (0.010)		
Adjusted R-squared		0.047		
D14: I try hard to always have some extra money for unexpected expenditures.		0.132*** (0.010)		
Adjusted R-squared		0.145		
Trust (N=1478)		Number of close friends		
Experiment		0.010 (0.027)		
Adjusted R-squared		0.063		
T24: 1st mover decision in hypothetical trust game		0.026 (0.028)		
Adjusted R-squared		0.063		
T16: People have only the best intentions		0.075*** (0.027)		
Adjusted R-squared		0.068		
T17: Willingness to trust neighbors		0.116*** (0.025)		
Adjusted R-squared		0.075		
T1: Willingness to trust		0.176*** (0.024)		
Adjusted R-squared		0.092		
Altruism (N=1646)		Social contribution	Pro-environment	Gender
Experiment		0.012 (0.011)	0.030 (0.021)	-0.002 (0.013)
Adjusted R-squared		0.221	0.365	0.032
A11: Hypothetical donation game		0.032*** (0.011)	0.069*** (0.021)	0.015 (0.012)
Adjusted R-squared		0.224	0.368	0.033
A10: Willingness to give to good causes		0.035*** (0.011)	0.067*** (0.022)	0.055*** (0.012)
Adjusted R-squared		0.225	0.368	0.043
A13: I am willing to donate time and money...		0.075*** (0.011)	0.090*** (0.023)	0.080*** (0.012)
Adjusted R-squared		0.241	0.371	0.055

Table C4: **Regression coefficients of preferences on individual behaviors.** In this table, we report OLS regressions of various preference measures on individual real-life behavior. The data is the pooled sample and contains only participants who answered control questions correctly. Items highlighted in blue are GPS survey items. We controlled for country fixed effects. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

D GPS and PSM Items and Measures

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	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 D1: **Overview: Behavioral Measures.** This table shows the behavioral measures for each preference and experiment.

Panel A:

GPS	Item description	Weight
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

Panel B:

PSM	Item description	Weight
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 D2: 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.

Measuring Economic Preferences with Behavioral Experiments and Surveys Across the Globe

– Online-Appendix –

Michael Kosfeld¹, Zahra Sharafi², Maíra Sontag González³, and Na Zou⁴

¹Goethe University Frankfurt, CEPR, CESifo, IZA, SAFE, and ZEF

²Frankfurt School of Finance and Management

³Goethe University Frankfurt

⁴University of Bath

January 6, 2026

A Replication Tables based on Full Sample

A.1 GPS Items, GPS Measure, and PSM Items

Pref.	Items	Germany	China	Iran	Kenya	US	Colombia	Rep. combined w FE
Risk taking	Staircase measure (R1)	0.216*** (0.054)	0.266*** (0.060) [0.148,0.385]	0.230*** (0.051) [0.129,0.331]	0.298*** (0.052) [0.196,0.400]	0.199** (0.062) [0.078,0.321]	0.332*** (0.053) [0.228,0.437]	0.273*** (0.025) [0.224,0.323]
	Willingness to take risks (R3)	0.241*** (0.051)	0.050 (0.060) [-0.069,0.169]	0.125*** (0.046) [0.034,0.216]	0.029 (0.054) [-0.078,0.136]	0.103* (0.056) [-0.007,0.213]	0.132*** (0.049) [0.036,0.229]	0.079*** (0.027) [0.027,0.131]
	Adjusted R-squared	0.140	0.080	0.084	0.088	0.057	0.154	0.148
Patience	Staircase measure (D1)	0.442*** (0.049)	0.417*** (0.051) [0.317,0.517]	0.219*** (0.054) [0.112,0.325]	0.190*** (0.065) [0.062,0.318]	0.369*** (0.054) [0.264,0.475]	0.323*** (0.053) [0.220,0.427]	0.343*** (0.024) [0.296,0.390]
	Willingness to wait (D3)	0.179*** (0.050)	0.130*** (0.052) [0.029,0.232]	0.180*** (0.052) [0.078,0.282]	0.056 (0.047) [-0.037,0.149]	0.240*** (0.048) [0.145,0.334]	0.064 (0.048) [-0.031,0.159]	0.112*** (0.020) [0.072,0.152]
	Adjusted R-squared	0.299	0.238	0.111	0.039	0.284	0.118	0.298
Trust	People have only the best intentions. (T16)	0.266*** (0.048)	0.166*** (0.050) [0.068,0.263]	0.182*** (0.050) [0.084,0.280]	0.191*** (0.053) [0.088,0.295]	0.166*** (0.053) [0.062,0.270]	0.135** (0.053) [0.030,0.240]	0.169*** (0.023) [0.123,0.214]
	Adjusted R-squared	0.070	0.025	0.031	0.034	0.025	0.016	0.059
Altruism	Hypothetical donation game (A11)	0.185*** (0.049)	0.243*** (0.052) [0.141,0.346]	0.262*** (0.060) [0.144,0.380]	0.175*** (0.059) [0.059,0.290]	0.279*** (0.054) [0.173,0.385]	0.178*** (0.054) [0.072,0.284]	0.230*** (0.026) [0.180,0.281]
	Share when it comes to good causes (A10)	0.321*** (0.044)	0.281*** (0.045) [0.192,0.370]	0.103* (0.054) [-0.002,0.209]	-0.053 (0.064) [-0.179,0.073]	0.216*** (0.039) [0.140,0.292]	0.139*** (0.045) [0.050,0.228]	0.155*** (0.024) [0.108,0.202]
	Adjusted R-squared	0.175	0.175	0.094	0.025	0.153	0.058	0.089
Pos. Reci- procity	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 (0.057) [-0.089,0.136]	0.263*** (0.048) [0.169,0.357]	0.102** (0.048) [0.007,0.197]	0.117*** (0.025) [0.067,0.166]
	Thank-you gift in the lost in the city scenario (PR9)	0.288*** (0.050)	0.201*** (0.047) [0.108,0.295]	0.163*** (0.051) [0.062,0.264]	0.036 (0.052) [-0.067,0.139]	0.151*** (0.052) [0.048,0.254]	0.221*** (0.053) [0.117,0.325]	0.149*** (0.025) [0.101,0.198]
	Adjusted R-squared	0.194	0.058	0.062	-0.003	0.107	0.063	0.071
Neg. Reci- procity	Punish someone who treats you unfairly (NR1)	N.A.	0.190*** (0.069) [0.054,0.325]	-0.044 (0.088) [-0.217,0.129]	0.096 (0.069) [-0.040,0.232]	0.266*** (0.084) [0.101,0.431]	0.235*** (0.066) [0.105,0.366]	0.155*** (0.035) [0.086,0.224]
	Punish someone who treats others unfairly (NR2)	N.A.	-0.089 (0.063) [-0.212,0.035]	0.074 (0.063) [-0.050,0.198]	-0.029 (0.064) [-0.155,0.098]	-0.002 (0.068) [-0.136,0.132]	-0.086 (0.070) [-0.224,0.051]	-0.024 (0.029) [-0.081,0.034]
	Take revenge at the first occasion (NR11)	N.A.	0.076 (0.067) [-0.055,0.207]	0.091 (0.085) [-0.075,0.258]	-0.018 (0.054) [-0.125,0.089]	-0.015 (0.071) [-0.155,0.124]	-0.080 (0.058) [-0.194,0.033]	0.008 (0.029) [-0.049,0.065]
	Adjusted R-squared	N.A.	0.038	0.004	-0.001	0.057	0.017	0.022
Number of observations		382 (360)	395	389	383	368	397	1932

Table A1: **Regression coefficients of GPS items - Full sample.** 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.

Pref.	Items	Germany	China	Iran	Kenya	US	Colombia	Rep. combined w FE
Risk taking	Staircase measure (R1) Willingness to take risks (R3)	0.353*** (0.052)	0.288*** (0.054) [0.183,0.394]	0.294*** (0.049) [0.197,0.391]	0.296*** (0.050) [0.197,0.394]	0.247*** (0.054) [0.141,0.353]	0.397*** (0.047) [0.306,0.489]	0.302*** (0.023) [0.257,0.347]
	Adjusted R-squared	0.124	0.081	0.084	0.085	0.058	0.156	0.148
Patience	Staircase measure (D1) Willingness to wait (D3)	0.539*** (0.040)	0.484*** (0.042) [0.401,0.567]	0.313*** (0.052) [0.210,0.416]	0.208*** (0.066) [0.078,0.338]	0.507*** (0.044) [0.420,0.594]	0.349*** (0.050) [0.252,0.446]	0.394*** (0.023) [0.349,0.439]
	Adjusted R-squared	0.289	0.232	0.096	0.041	0.255	0.120	0.294
Trust	People have only the best intentions. (T16)	0.266*** (0.048)	0.166*** (0.050) [0.068,0.263]	0.182*** (0.050) [0.084,0.280]	0.191*** (0.053) [0.088,0.295]	0.166*** (0.053) [0.062,0.270]	0.135** (0.053) [0.030,0.240]	0.169*** (0.023) [0.123,0.214]
	Adjusted R-squared	0.070	0.025	0.031	0.034	0.025	0.016	0.059
Altruism	Hypothetical donation game (A11) Share when it comes to good causes (A10)	0.321*** (0.048)	0.328*** (0.050) [0.231,0.426]	0.298*** (0.056) [0.188,0.409]	0.165*** (0.058) [0.051,0.280]	0.341*** (0.055) [0.234,0.449]	0.211*** (0.053) [0.107,0.314]	0.272*** (0.025) [0.222,0.321]
	Adjusted R-squared	0.091	0.106	0.087	0.025	0.114	0.042	0.071
Pos. Reci- procity	When someone does me a favor (PR11) Thank-you gift in the lost in the city scenario (PR9)	0.438*** (0.047)	0.237*** (0.047) [0.145,0.329]	0.259*** (0.044) [0.171,0.346]	0.048 (0.052) [-0.055,0.151]	0.317*** (0.049) [0.222,0.412]	0.253*** (0.044) [0.166,0.340]	0.207*** (0.024) [0.160,0.253]
	Adjusted R-squared	0.195	0.054	0.065	-0.000	0.098	0.062	0.071
Neg. Reci- procity	Punish someone who treats you unfairly (NR1) Punish someone who treats others unfairly (NR2) Take revenge at the first occasion (NR11)	NA	0.160*** (0.049) [0.064,0.257]	0.095* (0.050) [-0.003,0.194]	0.042 (0.053) [-0.062,0.146]	0.221*** (0.059) [0.106,0.337]	0.071 (0.051) [-0.030,0.172]	0.118*** (0.024) [0.071,0.166]
	Adjusted R-squared		0.023	0.007	-0.001	0.046	0.002	0.017
Number of observations		382 (360)	395	389	383	368	397	1932

Table A2: **Regression coefficients of GPS measures - Full sample.** 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. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Pref.	Items	Germany	China	Iran	Kenya	US	Colombia	Rep. combined w FE
Risk taking	Multiple price lists (R2)	0.276*** (0.057)	0.335*** (0.057)	0.261*** (0.055)	0.413*** (0.048)	0.310*** (0.063)	0.399*** (0.057)	0.359*** (0.025)
	Willingness to take risk (R3)	0.203*** (0.049)	0.016 (0.061)	0.092* (0.049)	0.007 (0.051)	0.063 (0.060)	0.101** (0.050)	0.044* (0.026)
			[-0.104,0.136]	[-0.004,0.188]	[-0.094,0.107]	[-0.055,0.180]	[0.003,0.199]	[-0.007,0.094]
	Adjusted R-squared	0.162	0.112	0.093	0.167	0.107	0.195	0.193
Patience	Multiple price lists (D2)	0.485*** (0.052)	0.470*** (0.049)	0.192*** (0.057)	0.186*** (0.062)	0.469*** (0.051)	0.383*** (0.051)	0.374*** (0.024)
	Willingness to wait in comparison with others? (D4)	0.171*** (0.050)	0.076 (0.049)	0.195*** (0.054)	-0.006 (0.044)	0.131*** (0.049)	0.073 (0.048)	0.079*** (0.020)
			[-0.021,0.173]	[0.089,0.302]	[-0.094,0.081]	[0.034,0.227]	[-0.022,0.167]	[0.040,0.118]
	Adjusted R-squared	0.340	0.256	0.101	0.029	0.289	0.163	0.305
Trust	First-mover behavior in trust game (T24)	0.629*** (0.043)	0.570*** (0.051)	0.471*** (0.057)	0.310*** (0.054)	0.566*** (0.045)	0.301*** (0.052)	0.438*** (0.024)
	People have only the best intentions. (T16)	0.133*** (0.038)	0.057 (0.040)	0.108** (0.046)	0.162*** (0.049)	0.083* (0.043)	0.094* (0.052)	0.107*** (0.021)
			[-0.021,0.135]	[0.017,0.199]	[0.066,0.258]	[-0.002,0.168]	[-0.008,0.197]	[0.065,0.148]
	Adjusted R-squared	0.452	0.337	0.245	0.127	0.338	0.103	0.240
Altruism	Hypothetical donation game (A11)	0.185*** (0.049)	0.243*** (0.052)	0.262*** (0.060)	0.175*** (0.059)	0.279*** (0.054)	0.178*** (0.054)	0.230*** (0.026)
	Share when it comes to good causes. (A10)	0.321*** (0.044)	0.281*** (0.045)	0.103* (0.054)	-0.053 (0.064)	0.216*** (0.039)	0.139*** (0.045)	0.155*** (0.024)
			[0.192,0.370]	[-0.002,0.209]	[-0.179,0.073]	[0.140,0.292]	[0.050,0.228]	[0.108,0.202]
	Adjusted R-squared	0.175	0.175	0.094	0.025	0.153	0.058	0.089
Pos. Reci- procity	Second-mover behavior in trust game (PR16)	0.486*** (0.049)	0.506*** (0.045)	0.399*** (0.049)	0.257*** (0.060)	0.416*** (0.074)	0.158*** (0.053)	0.341*** (0.028)
	Thank-you gift in the lost in the city scenario (PR9)	0.164*** (0.049)	0.069* (0.042)	0.098* (0.050)	-0.028 (0.052)	0.111** (0.055)	0.205*** (0.053)	0.091*** (0.024)
			[-0.013,0.151]	[-0.000,0.196]	[-0.130,0.074]	[0.004,0.219]	[0.101,0.308]	[0.043,0.139]
	Adjusted R-squared	0.329	0.279	0.185	0.058	0.204	0.076	0.157
Neg. Reci- procity	Minimum acceptable offer (NR10)	0.340*** (0.058)	0.226*** (0.046)	0.121** (0.055)	0.054 (0.061)	0.273*** (0.051)	0.054 (0.049)	0.130*** (0.024)
			[0.135,0.317]	[0.014,0.228]	[-0.065,0.174]	[0.171,0.374]	[-0.041,0.150]	[0.084,0.177]
	Adjusted R-squared	0.115	0.049	0.012	0.000	0.072	0.000	0.020
Number of observations		382 (360)	395	389	383	368	397	1932

Table A3: **Regression coefficients of PSM items - Full sample.** 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. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

A.2 PCA, Survey Modules, Cross-Preference Correlations

Preference	Principal components	Corresponding survey questions	Fraction of Expl. Var.
A: Risk taking	Willingness to take risks	R3-R8, R12, R15-R16	0.302
	<i>Quantitative measures for risk</i>	Exp. incen. measure, R1-R2, R16	0.123
	Risk in financial investments	R13-R15	0.096
	Risk in daily activities	R9-R11	0.093
B: Patience	Willingness to wait in financial situations	D5-D6, D8-D11, D14-D15	0.140
	Willingness to wait	D3-D5, D7-D8, D17	0.127
	<i>Quantitative measures for patience</i>	Exp. incen. measure, D1, D2	0.117
	Not keeping resolutions	D11, D14, D16-D18	0.106
	Compensation for giving up a vacation trip	D7-D8, D12-D13	0.054
C: Trust	Willingness to trust	T1-T8, T13-T14, T17	0.140
	Most people are fair	T20, T21	0.085
	Building trust in other people and their intentions	T10-T13, T16	0.076
	Most people act in their own best interest	T22, T23	0.071
	<i>Quantitative measures for trust</i>	Exp. incen. measure, T24	0.059
	Willingness to rely on strangers	T15, T18-T19	0.050
	Willingness to trust neighbor	T4, T8-T9	0.047
D: Altruism	Willingness to share with people	A1-A10, A14	0.285
	Willingness to do good if I expect it returns	A12, A15-A17, A20	0.103
	Be an unselfish person	A18, A19, A21	0.075
	<i>Quantitative measures for altruism</i>	Exp. incen. measure, A10-A11, A13, A17	0.075
	Hours of volunteering	A17, A22-A23	0.060
E: Pos. reciprocity	Willingness to return a favor	PR1-PR8	0.289
	I help someone who has helped me before	PR11-PR15	0.177
	Thank-you gift to return a favor	PR9-PR10, PR16	0.101
	<i>Quantitative measures for positive reciprocity</i>	Exp. incen. measure, PR14, PR16	0.080
F: Neg. reciprocity	Willingness to punish people with unfair behavior	NR2-NR9	0.232
	Willingness to punish people who treated you unfairly	NR1, NR11-NR14, NR16, NR18	0.208
	Dislike being taken for a fool and pushed around	NR15, NR17, NR19-NR20, NR22	0.101
	<i>Quantitative measures for negative reciprocity</i>	Experimental incen. measure, NR10	0.050
	Fairness in hypothetical scenario	NR17, NR21	0.046

Table A4: **Principal components for each preference - Full Sample.** 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.

Preference	Module	Item	Countries	Adj R^2	MSE test
Risk taking	Single-item	Quant - R2: Multiple price lists	CN; CO; DE; IR; KE; US	0.190	0.859
		Quant - R1: Switching row from staircase measure	CN; CO; DE; IR; KE; US	0.143	0.923
		Qual - R16: You win 100,000 EUR in a lottery. How much would you invest in the risky yet profitable lottery?	CN; CO; DE; IR; KE; US	0.084	1.020
	Multiple-item (Random Forest)	R1: Switching row from staircase measure Quant - R3: Willingness to take risks R16: You win 100K EUR in a lottery. How much of it would you invest in the risky yet profitable lottery?		0.154	1.162
Patience	Single-item	Quant - D2: Multiple price lists (D2)	CN; CO; DE; IR; KE; US	0.365	0.713
		Quant - D1: Switching row from staircase measure	CN; CO; DE; IR; KE; US	0.350	0.759
		Qual - D3: Willingness to wait	CN; CO; DE; IR; US	0.279	0.988
	Multiple-item (Random Forest)	D1: Switching row from staircase procedure D3: Willingness to wait D18: I am a person who often does not keep their own good resolutions		0.361	0.764
Trust	Single-item	Quant - T24: First-mover behavior in trust game	CN; CO; DE; IR; KE; US	0.211	0.867
		Qual - T1: Willingness to trust	CN; CO; DE; IR; KE; US	0.063	0.997
		Qual - T17: In general, one can trust other people.	CN; CO; DE; IR; KE; US	0.061	0.985
	Multiple-item (Random Forest)	T16: I assume that people have only the best intentions. T17: Most people only act in their own best interest T24: 1st mover decision in hypothetical trust game		0.230	0.838
Altruism	Single-item	Quant - A11: Hypothetical donation game	CN; CO; DE; IR; KE; US	0.094	0.916
		Qual - A13: I am willing to donate time and money to charity, even if I do not profit from that directly	CN; CO; DE; IR; US	0.100	0.907
		Qual - A10: Willingness to give to good causes	CN; CO; DE; IR; US	0.083	0.933
	Multiple-item (Random Forest)	A10: Willingness to give to good causes A11: Hypothetical donation game A13: I am willing to donate time and money to charity, even if I do not profit from that directly		0.136	1.091
Pos. Reciprocity	Single-item	Quant - PR16: 2nd mover behavior in hypothetical trust game	CN; CO; DE; IR; KE; US	0.169	0.945
		Qual - PR9: Thank-you gift in the lost in the city scenario	CN; CO; DE; IR; US	0.063	1.027
		Qual - PR11: When someone does me a favor	CN; CO; DE; IR; US	0.059	1.011
	Multiple-item (Random Forest)	PR9: Thank-you gift in the lost in the city scenario PR11: When someone does me a favor PR16: 2nd mover behavior in hypothetical trust game		0.182	0.897
Neg. Reciprocity	Single-item	NR10: Minimum acceptable offer in hypothetical Ultimatum Game	CN; DE; IR; US	0.306	1.024
		NR12: When someone puts me into a difficult situation, I will do the same to them	CN; DE; IR; US	0.304	1.044
	Multiple-item (Random Forest)	NR19: I absolutely dislike being the fool. NR21: Hypothetical scenario: cutting wages NR22: It is important to me to be respected by others.		0.296	0.859

Table A5: Alternative Preference Modules - Full sample. 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.

Preference	Items	Countries	Summary
Risk	1. Multiple price lists	DE, CN, IR, KE, US, CO All_Rep, All	R2
	2. Imagine you win 100.000 Euro in a lottery. You can participate in the lottery using the whole amount you won or only a part of it	CN, IR, All_Rep, All	R16
	3. Willingness to take risks	DE	R3
	4. Willingness to take risks in financial investments	CO	R6
	5. Willingness to take risks in important decisions in life	KE	R7
	6. Willingness to take risks in traffic	US	R10
Patience	1. Multiple price lists	DE, CN, US, CO All_Rep, All	D2
	2. Staircase measure	IR, KE, All_Rep	D1
	3. Willingness to wait?	IR, All	D3
	4. Willingness to wait in comparison with others?	DE	D4
	5. Willingness to wait assessed by others	CO	D5
	6. Willingness to wait wrt. financial investments	US	D6
	7. I am a person who often does not keep their own good resolutions	CN	D18
	8. How much money do you save per month?	KE	D19
Trust	1. Hypothetical first-mover behavior	DE, CN, IR, KE, CO, US, All_Rep, All	T24
	2. People have only the best intentions	DE, KE, All	T16
	3. Willingness to trust people in your professional environment	All_Rep	T6
	4. Willingness to trust compared to others	US	T2
	5. Willingness to trust people in your hometown	CO	T4
	6. Willingness to trust people in your neighborhood	CN	T8
	7. Most people would take advantage of you when they have the chance	IR	T20
Altruism	1. Hypothetical donation game	DE, CN, IR, KE, CO, US, All_Rep, All	A11
	2. I am willing to donate time and money to charity, even if I do not profit from that directly.	CN, IR, CO, All_Rep; All	A13
	3. Share when it comes to good causes.	DE, US	A10
	4. I am a person who would give their shirt off their back to help others	KE	A18
Pos. Reciprocity	1. Hypothetical second-mover behavior	DE, CN, IR, KE, US, All_Rep, All	PR16
	2. Willingness to return a favor when it comes to strangers	CN, All_Rep; All	PR7
	3. Thank-you gift in the lost in the city scenario	DE, CO	PR9
	4. I do not like the feeling of owing something to someone	IR, CO	PR14
	5. I am willing to incur costs to help someone who has helped me before	US	PR13
Neg. Reciprocity	1. Minimum acceptable offer in ultimatum game	DE, CN, IR, US, All_Rep, All	NR10
	2. Willingness to punish someone who treats you unfairly	All_Rep, CO, US	NR1
	3. Willing to punish unfair behavior even if this is costly	DE	NR1/2
	4. Willingness to punish compared with others	CO	NR3
	5. Willingness to punish your circle of friends	KE	NR6
	6. Willingness to punish unfair behavior when it comes to your professional environment.	IR	NR7
	7. Willingness to punish unfair behavior when it comes to your neighborhood	KE	NR9
	8. When someone puts me into a difficult situation, I will do the same to them	All	NR12
	9. If someone harms me on purpose, I will try to give that person a taste of their own medicine	CN	NR14
Number of observations: DE: 382(360); CN: 395; IR: 389; KE: 383, CO: 397, US:368, All (replication combined): 1167, All: 1549 (1527)			

Table A6: Best combination of survey items - Full Sample. This table shows the best combination of survey items based on the item selection procedure in ?. Blue-colored items are GPS items.

Qualitative Self-Assessment	Incentivized Elicitations					
	<i>Risk</i>	<i>Time</i>	<i>Altruism</i>	<i>Trust</i>	<i>Pos. Reciprocity</i>	<i>Neg. Reciprocity</i>
Risk	0.187*** -0.024	0.017 -0.019	0.073*** -0.023	0.136*** -0.023	0.043* -0.025	0.050*** -0.018
Patience	0.039 -0.024	0.273*** -0.02	0.060** -0.024	0.056** -0.023	0.090*** -0.025	-0.013 -0.019
Trust	0.022 -0.022	-0.004 -0.019	0.149*** -0.021	0.187*** -0.021	0.131*** -0.022	-0.027 -0.018
Altruism	-0.004 -0.021	0.012 -0.019	0.249*** -0.021	0.052** -0.022	0.145*** -0.022	-0.045** -0.02
Pos. Reciprocity	0.006 -0.024	-0.016 -0.019	0.148*** -0.022	0.171*** -0.022	0.207*** -0.022	-0.068*** -0.02
Neg. Reciprocity	0.018 -0.026	-0.037* -0.022	-0.114*** -0.024	-0.052** -0.026	-0.104*** -0.026	0.134*** -0.023

Table A7: **Correlations between Qualitative Self-Assessments and Incentivized Elicitations - Full Sample.** ***, **, * denote statistical significance at the 1%, 5%, and 10% level.

B Experiments and Survey Questions

B.1 Experiments (incl. Control Questions)

Original instructions in English are available at:

<https://nextcloud.flex.uni-frankfurt.de/index.php/s/ZHSRXyiFo3GcEfM>

Risk taking

Two multiple price lists are used 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?

Patience

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 ?. 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 to receive 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 the Red Cross and Greenpeace, but replaced German charitable organizations with domestic charitable organizations in respective countries. Since donations were made by the experimenter on behalf of participants, we made it clear that all participants can access 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 questions. 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 ?. 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 (?). 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?

B.2 Survey items

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

Patience

- **(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.

B.3 Excluded questions from the original survey battery in ?

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?