CRESTCENTER FOR RESEARCH IN ECONOMICS AND STATISTICS

WORKING PAPER SERIES

Should we trust measures of trust?

Héloise Cloléry, Guillaume Hollard,
Fabien Perez, Inès Picard

N°13/July 2022

CREST
Center for Research in Economics and Statistics
UMR 9194

5 Avenue Henry Le Chatelier
TSA 96642
91764 Palaiseau Cedex
FRANCE

Phone: +33 (0)1 70 26 67 00
Email: info@crest.science
https://crest.science/

Should we trust measures of trust?*

Héloise Cloléry[†], Guillaume Hollard[‡], Fabien Perez[§], Inès Picard[¶]

July 8, 2022

Abstract

Trust is an important economic variable that may however be subject to measurement error, leading to econometric issues such as attenuation bias or spurious correlations. We use a test/retest protocol to assess the measurement error in the two main tasks that are used to elicit trust, namely survey questions and experimental games. We find that trust measures based on the trust game entail substantial measurement error (with up to 15% of noise), while there is virtually no noise in stated trust measures. Given the specificity of our subject pool (students in a top Engineering school) and the short period of time between the test and the retest, we consider these percentages of noise as lower bounds. We also provide a sub-group analysis based on measures of cognitive ability and effort. We find substantial heterogeneity across sub-groups in trust-game behavior, but none for the survey questions. We finally discuss which measure of trust should be used, and the estimation strategies that can be applied to limit the effect of measurement error.

Keywords: Trust; Trust Game; Measurement Error; ORIV.

JEL Classification: C18; C26; C91; D91

*We would like to thank Wael Bousselmi, Bertrand Garbinti, Lucas Girard, Yves Le Yaouanq and Radu Vranceanu for their helpful comments. We are also grateful to participants at the 2021 ESA Global Online Conference, the 2021 ASFEE conference, the 2022 JMA conference and the 2022 LAGV conference for their suggestions.

[†]CREST-Ecole polytechnique, IP Paris, heloise.clolery@polytechnique.edu. 5 Av. Le Chatelier, 91120 Palaiseau, FRANCE.

[‡]CREST-Ecole polytechnique, IP Paris and CNRS - guillaume.hollard@polytechnique.edu.

[§]CREST-Ensae, IP Paris - fabien.perez@ensae.fr.

[¶]CREST-Genes, IP Paris ines.picard@ensae.fr

1. Introduction

Subjective characteristics, such as risk aversion, discount rate, or trust, are frequently used in regressions and shown to be statistically significant. However, Gillen et al. (2019) warn researchers that subjective characteristics may be prone to large measurement error. As a result, regressions including subjective characteristics may have biased coefficients and exhibit spurious correlations. Researchers should thus be cautious in their estimation strategy for variables that include a large amount of noise. Trust is an important economic variable, yet we still know little about the extent of noise in its measures. How serious is the measurement error issue for elicited measures of trust? Different trust elicitation methods have been used, like surveys and incentivized experiments, but with no comparison in terms of errors generated by these techniques.

To gauge measurement error for elicited measures of trust, we here use a test/retest procedure. The principle of test/retest is to have subjects performing the same elicitation task twice within a short period of time. Differences in subjects' answers between the two repetitions are one realization of the change in errors. Noise corresponds to the discrepancy in changes across respondents. Formally, we compute the ratio of the variance in noise across all individuals over the variance in the elicited measure. This noise ratio is computed for the two most popular elicitation methods: surveys and trust games.

Our results are twofold. First, we find that survey questions are remarkably stable in our test/retest design. In contrast, the behavior in the trust game is noisier, with the amounts returned being less noisy than the amounts sent. Compared to similar exercises run on measures of risk attitudes, we find that trust measures are less noisy than risk-attitude measures. Second, we identify two important drivers of the amount of noise in the trust game: cognitive ability and individual effort (or attention). Subjects with greater cognitive ability produce measures that are very stable between tasks. Equally, subjects who put more effort into their work are also less prone to noisy behavior. One interpretation is that noise can be avoided by subjects who either apply a deliberate strategy in the trust game or who focus sufficiently on their work to deliver less noisy answers. In sharp contrast, the measures from survey questions incorporate very little noise, irrespective of the characteristics that appear to lie behind the amount of noise in game-based measures.

Which measure of trust should be used in practice? Researchers who are primarily concerned with the amount of noise in their trust measures may consider using survey questions. Those who attach great importance to incentives should anticipate that running a trust game, especially among the non-student population, will lead to substantial measurement error. A strategy to handle measurement error is then required. The ORIV method, proposed by Gillen et al. (2019), is an effective strategy to cope with measurement error, but at the price of repeating the same measures twice. We present and discuss alternative strategies to handle measurement error which may be implemented based on a single measure.

Last, we discuss how proper controls on measurement error may solve two key questions in the literature on trust. First, can the absence of correlation, sometimes reported[1], between stated and incentivized measures of trust be attributed to measurement error? If we consider the noisiest sub-group in our sample, the noise ratio reaches 37%. Even if the noise ratio were to reach 50% in some studies, correcting for noise would *at best* double observed correlations. Even in this extreme scenario, correlations close to zero are thus likely to remain insignificant once measurement error is accounted for. Measurement error on trust is therefore unlikely to explain the (sometimes) observed lack of correlation. Second, to what extent trust may be a proxy of other variables like risk attitude? We find little and non-significant correlation between trust and risk. Hence, handling measurement error is unlikely to push observed correlations up. We find no evidence that trust should be considered as a by-product of risk-attitude.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 introduces the experimental protocol. Section 4 presents the results. We provide a discussion in Section 5, and Section 6 concludes.

2. Review

2.1. The importance of trust in economics

The economic importance of trust cannot be overstated. To give just one example, in a world of incomplete contracts, an element of trust is required to cope with unforeseen contingencies. In particular, Arrow (1972) and Fukuyama (1995) underlines the importance of trust in all commercial relationships. As such, most economic relations and organizations

[1]Glaeser et al. (2000) found no correlation between various measures of trust, which initiated a still ongoing debate on the question. See section 2.2. for a literature review.

would benefit from greater trust (La Porta et al., 1997). Indeed, trust-induced cooperative behavior is beneficial for international trade, and all the more when the system of laws governing import is not sufficiently good for the exporting country (Yu et al., 2015). In addition, theoretical and empirical works found sizeable effects of trust on GDP and worldwide growth. For example, Knack and Keefer (1997) report positive correlation between trust and growth rate, as confirm by subsequent papers (Zak and Knack, 2001; Algan and Cahuc, 2010; Horváth, 2013). More than just affecting the GDP, trust affects macroeconomic stability (Sangnier, 2013) and the policy implementation (Algan et al. (2016) found a twin-peak relation between trust and welfare state).

2.2. What is trust and (how) can it be measured?

In economics, trust is a behavior in which a trustor places resources at the disposal of a trustee. The trustor is expecting the trustee to undertake specific actions but cannot force in any way the trustee to do so (e.g. there is no possible punishment). Berg et al. (1995) introduce the trust game which fairly describes such a situation. So, the amounts sent in the trust game are often used as a measure of trust. An alternative method to measure trust is to rely on survey questions such as "Do you think that most people can be trusted?". Such questions have been widely used, among others in the World Value Survey. Whatever the measure used, substantial between-subjects variations are found (see Johnson and Mislin (2011) for a review focusing on the trust game, and (Cooper and Kagel, 2016) for a review on other-regarding preferences).

Both methods are expected to capture a "core component" that would, implicitly, define trust as the common factor to all measures of trust. The existence of a core component or latent variable is not granted, however. Indeed, Glaeser et al. (2000), Lazzarini et al. (2003) and Ashraf et al. (2004) find no correlation between stated trust and trust in games, but do find correlations between stated trust and trustworthiness[2]. On the contrary, Fehr et al. (2003) and Bellemare and Kröger (2007) find that stated trust correlates with trust in experimental games but not with trustworthiness. Veszteg et al. (2015) observe that stated trust is correlated with both trust and trustworthiness behavior in the trust game. Falk et al. (2016) show that stated trust can predict trust in games. Thöni et al. (2012) find that trust in

[2]Trustworthiness measure corresponds to the second stage of the trust game with the amount sent back by the trustee to the trustor

games is linked to other-regarding behavior while Sapienza et al. (2013) and Banerjee (2018) find on the contrary that it is linked to beliefs about trustworthiness. At this point in time, there is no clear consensus on what may drive correlations among different measures of trust. For instance, Aksoy et al. (2018) show that some aspects of the trust game are responsible for large changes in the correlation. Generally speaking, this “correlation puzzle” is addressed by trying to better control for relevant covariates affecting each measure. In contrast with these approaches, we here explore the possibility that measurement error affecting elicited values are in part responsible for the correlation puzzle.

2.3. Measurement Error

We know at least from the work of Spearman (1904) that correlations and coefficients in OLS regressions can be biased if variables are measured with error. Spearman is the first to refer to the *attenuation bias*: the correlation between variables measured with error is lower in absolute value. More than a century later, Gillen et al. (2019) deal with the specific problem of measurement error in laboratory experiments and propose an instrumental approach denoted as ORIV (Obviously Related Instrumental Variables). Gillen et al. argue that errors in measurement are ubiquitous in experiments but rarely taken into account. As a matter of fact, when we ask subjects to perform twice the same task in a short period of time, we often obtain different choices. Gauging measurement error in experimental tasks is very important to assess the magnitude of the attenuation bias in regressions or correlations. Furthermore, when estimating multivariate OLS regressions, the coefficient of variables measured without error can be biased if one variable included in the regression is measured with error. Gillen et al. (2019) show that, in some studies, the gender variable wrongly appears significant due to measurement error in risk-attitude. Perez et al. (2021) build on their work and quantify the amount of noise in 4 different risk-aversion tasks using 16 datasets. Perez et al. (2021) find noise ratios between 35% and 60% for risk measures. Corresponding figures are not yet available for measures of trust.

3. Experimental Design

The purpose of our design is to gauge the amount of noise in elicited measures of trust and risk. We choose a test/retest procedure that consists in subjects performing the same tasks twice within a short period of time (about 15 minutes in our experiment). The great advantage of the test/retest design is that it provides very simple non-parametric measures of noise. Participants face the following sequence of tasks (see the detailed descriptions below): (1) Holt and Laury risk measure,[3] (2) Trust Game as the Sender[4], (3) Trust Game as the Receiver[5], (4) Holt and Laury measure, (5) Stated Trust, (6) Distracting Tasks, (7) Trust Game as the Sender[6], (8) Trust Game as the Receiver[7], (9) Stated Trust. We chose the most widely-used risk and trust measures in the literature. For each incentivized task, the outcomes were shown to the participants only at the very end of the experiment. The payoffs were expressed in coins: 10 coins represented €5.

3.1. The Trust Game

We use the standard trust game as proposed by Berg et al. (1995) (also known as the investment game). The trust game is, by far, the most popular incentivized measure of trust and trustworthiness (for instance, Van Den Akker et al. (2020) find 167 papers that have used this trust game). The Trust game is a two-stage two-player game with a sender and a receiver. In our experiment, both the sender and the receiver have an initial endowment of 10 ECU (i.e. €5). In the first stage, the sender chooses how much of her 10 ECU endowment she wants to send to the receiver (the amounts have to be integers). The amount sent is denoted as X and is the measure of trust. The receiver then obtains X times 3 and chooses Y, the amount she wants to send back to the sender (Y has to be between 0 and $3X$). Y is the measure of trustworthiness. We elicit trustworthiness using the strategy method: subjects are asked to provide a value of Y for the 10 possible values of X corresponding to the integers between 1 and 10. Subjects play each role twice, both times against a random player.

[3]Subjects participated in two back-to-back experiments. These were independent (i.e. presented by different experimenters, for an unrelated purpose, were paid independently etc.). The experiment we discuss here is the second one. The first risk-attitude measure was elicited during the other experiment. Having two experiments with the same pool of subjects allows us to compare risk attitudes across experiments.

[4]against a randomly matched participant

[5]against the same participant as in task (2)

[6]against a different randomly selected participant than in tasks (2) and (3)

[7]against the same participant as in task (7)

3.2. Stated Trust

A popular way to elicit trust is via a variety of survey questions. We here consider the one used in, among others, the World Values Survey (WVS) and the US General Social Survey (GSS). In 2013, Sapienza et al. (2013) counted more than 500 studies using the WVS question: "*Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?*".[8]

The response scale differs across surveys. For instance, the WVS and the GSS propose only two possible answers ("most people can be trusted" and "you can't be too careful"). We here use a 0 to 10 scale, where 0 means "you can't be too careful" and 10 "most people can be trusted". We believe this scale to be more appropriate in detecting variations in test/retest analysis; it has successfully been used in, among others, the European Social Survey. Subjects answered this question twice during the course of the experiment.

3.3. Measure of risk-attitudes: Holt and Laury

Holt and Laury (2002) proposed a measure of risk attitudes via multiple binary choices between lotteries. The Holt-Laury method (HL) is a standard measure of risk-attitudes, and its test/retest properties have been analyzed in a number of contributions (see Perez et al., 2021, for an overview). We compare choices in two Holt-Laury tasks, one from a first experiment and the second from the present experiment. The two tasks are not presented in exactly the same way but involve mathematically-equivalent choices (see the Appendix for screenshots). The amounts involved are roughly equal to those used in the original HL experiment.

3.4. Distracting tasks

Test/retest designs use distracting tasks to avoid making the repetition of identical tasks too salient. We here use a task in which subjects have to count the number of ones in as many matrices as possible over a period of five minutes. We also included six incentivized questions that are variants of the cognitive reflection test (CRT).[9] We increase the cognitive burden by asking subjects to memorize a seven-digit number (1429587) that they have to

[8]The question is slightly differently formulated in the WVS.

[9]The original CRT was introduced in Frederick (2005): we use variants that correspond to the three first questions in Finucane and Gullion (2010) and questions 4 to 6 from Toplak et al. (2014).

report once the distracting tasks are finished. All tasks are incentivized (reporting correctly the seven-digit number is paid 6 coins, each correct matrix count is rewarded by 2 coins with a penalty of -1 coin for incorrect answers, and each correct answer to a CRT question yields 2 coins).

4. Results

The experiment took place at CREST Experimental Lab using the O-Tree platform (Chen et al., 2016). We used ORSEE (Greiner, 2004) to recruit 155 students for participation in the experiment. The experiment took place in the Fall of 2021.

4.1. Descriptive statistics

Table 1 shows the descriptive statistics for the main variables that appear in our experiment. We carried out Mann-Whitney and Kolmogorov-Smirnov tests to check that the repetitions of each task produce the same distribution: this is the case in our data for all variables apart from the Holt and Laury task. We discuss this in Appendix B. Our mean values for the amounts sent and the fractions returned in the trust game are similar to those in previous work. We also obtain a similar distribution of transfers, with a large fraction of participants sending either 0 or 10 coins, and a peak at 5.[10] The distributions of our different variables are discussed in Appendix A.

[10] See Capra et al. (2008) for a review of experimental results.

Table 1: Descriptive statistics

	No. Obs.	Mean	SD	Min	Max	MW	KS
Amount sent 1	155	4.42	3.63	0	10	.942	1
Amount sent 2	155	4.39	3.68	0	10		
Stated Trust 1	155	4.85	2.55	0	10	.977	1
Stated Trust 2	155	4.85	2.55	0	10		
HL1	155	5.37	1.69	0	9	.003	.049
HL2	155	5.95	1.79	0	10		
Fraction Returned 1	155	.32	.24	0	1	.902	.956
Fraction Returned 2	155	.32	.24	0	1		
CRT	155	4.49	1.61	0	6		
Grid Score	155	13.1	5.53	0	24		
Memorize	155	.93	.26	0	1		

Notes: HL1 (HL2) is the number of safe choices in the first (second) Holt and Laury task. CRT is the number of correct answers to the six CRT-like questions. The Grid Score is twice the number of correct counts of ones minus the number of incorrect counts of ones in the counting task. In the case of negative payoffs, the Grid Score is set to 0. Memorize equals 1 if the subject correctly memorized the 7-digit number.

4.2. Measurement Error

We estimate measurement error in each task using a non-parametric method set out below. We note that rounding issues (i.e. the fact that respondents report integers) is not taken into account. This issue is addressed in Perez et al. (2021), which also proposes non-parametric estimation. We provide similar parametric estimations in Appendix D.[11]

We assume that we observe two noisy measures for each variable x:[12]

$$x_1 = x^* + \epsilon_1 \quad \text{and} \quad x_2 = x^* + \epsilon_2$$

We assume independence between the errors ϵ and the true parameter x^*, and that the

[11] Perez et al. (2021) show empirically that parametric and non-parametric methods yield similar estimates of measurement error in four different risk-aversion tasks, using 16 datasets.

[12] Adding a systematic error to the Holt and Laury measures does not change the estimations. See Appendix B and Appendix D for details.

ϵ are independent and identically-distributed across repetitions.[13]

Then $Var(x_1 - x_2) = Var(\epsilon_1 - \epsilon_2) = 2Var(\epsilon)$, as ϵ_1 and ϵ_2 are assumed to be independent.

$$Var(\epsilon) = \frac{Var(x_1 - x_2)}{2}$$

We can therefore estimate the variance of the noise in measurement using the empirical variances:

$$\widehat{\sigma_\epsilon}^2 = \frac{\widehat{Var}(x_1 - x_2)}{2}$$

The variance of the measure can be estimated by the mean of the empirical variances of x_1 and x_2

$$\widehat{\sigma_m}^2 = \frac{\widehat{Var}(x_1) + \widehat{Var}(x_2)}{2}$$

We are interested in the noise ratio R, defined as the part of the measure's variance that reflects measurement error (noise):

$$R = \frac{\sigma_\epsilon^2}{\sigma_m^2} \qquad \text{or equivalently} \qquad R = 1 - Corr(x_1, x_2)$$

We can estimate this ratio by:

$$\widehat{R} = \frac{\widehat{\sigma_\epsilon}^2}{\widehat{\sigma_m}^2} \qquad \text{or simply by} \qquad \widehat{R} = 1 - \widehat{Corr}(x_1, x_2)$$

[13] We discuss the relaxation of these independence assumption in appendix

Table 2: Measurement error

	$\widehat{\sigma_m}^2$	$\widehat{\sigma_\epsilon}^2$	Noise Ratio $\widehat{R}$
Stated Trust	6.5	.32	**4.89%**
Fraction Returned	.057	.0047	**8.24%**
Amount Sent	13.37	2.00	**14.9%**
HL	3.02	1.36	**45.1%**

Notes: For each elicited measure, this table shows the estimated variance $\widehat{\sigma_m}^2$, the estimated variance of the error term $\widehat{\sigma_\epsilon}^2$, and the estimated noise ratio $\widehat{R}$.

We can see in Table 2 that the amount of noise varies greatly across measures. The noise ratio in the risk measure reaches a figure of 45%, very much in line with the ratio found in other datasets (see Perez et al., 2021). By way of contrast, the amounts sent in the trust game appear much less noisy, with a noise ratio of 15%. Trustworthiness, as measured by the average fraction of money that is returned in the trust game, appears to entail only little noise,[14] while at the extreme, stated trust is remarkably stable and (almost) immune to measurement error.

These estimations of measurement error (in particular in the trust tasks) should be interpreted as lower bounds of the noise in typical experiments for three reasons. First, the amount of time elapsed between the test/retest measures was quite short. Second, subjects were in a well-equipped experimental laboratory, synonymous with excellent material conditions. Last, our subjects are students in top Engineering schools, and are therefore presumably less prone to noisy behavior.

4.3. Sub-group analysis - What drives the noise in experimental measures?

The previous section provides estimates of the *average* amount of noise. We here, in contrast, explore how the amount of noise varies across specific sub-groups. Our sample

[14]The fact that we average measures may explain the smaller noise figure for this measure: see Appendix C.

is obviously limited in size, and covers a particular population (students in Engineering schools) that restricts its external validity. Rather than conducting a systematic review of all of the potential covariates, we focus on two potential drivers of noise identified in previous work: cognitive ability and effort.[15] Cognitive ability is expected to play a more important role in tasks involving strategic uncertainty, and has been found to be correlated with noise (greater cognitive ability leads to less noise). Effort, which can also be considered as a proxy for attention, is also important as poor attention may produce more noise. Furthermore, test/retest designs involving strategic uncertainty are affected by two key variables: (1) The ability to identify a strategy in games, known as strategic ability (e.g. the ability to form beliefs or calculate a best-response), and (2) The level of attention, which makes it more likely to *remember* previous choices. We here measure cognitive ability using the six CRT-questions measure, and effort from the score obtained in the real-effort task used as a distractor. We use a median split rule to classify subjects into four categories over two dimensions: cognitive ability and effort/attention. Subjects with a grid score higher (lower) than 12.5 are labeled *High Effort* (*Low Effort*); subjects who have a CRT score higher (lower) than 4.5 are labeled *High CRT* (*Low CRT*). The following table lists the values for each case and for each variable. Effort and CRT are relatively uncorrelated, so that the cells contain a similar number of observations. We carry out the same analysis for risk-attitudes to provide a benchmark.

[15] For instance, Amador-Hidalgo et al. (2021) find that noise and inconsistent choices in risky tasks are negatively correlated with cognitive ability. In Anderson and Mellor (2009), when subjects are consistent in their answers to survey questions - which they interpret as a measure of their effort or comprehension in the experiment - they also exhibit more stable risk preferences.

Table 3: Estimation of the Noise Ratios in Trust Behavior by Sub-groups

	Low CRT		High CRT		Total	
Low Effort	Incentivized Trust:	37.4%	Incentivized Trust:	17.9%	Incentivized Trust:	24.3%
		[19.5%,65.7%]		[9.8%,31.9%]		[15.7%,36.6%]
	Trustworthiness:	5.3%	Trustworthiness:	15.2%	Trustworthiness:	10.2%
		[2.5%,11%]		[8.2%,27.3%]		[6.4%,16.1%]
	Stated Trust:	3.7%	Stated Trust:	6.3%	Stated Trust:	5.1%
		[1.7%,7.7%]		[3.3%,11.8%]		[3.2%,8.1%]
	Risk:	92.6%	Risk:	39.2%	Risk:	61.2%
		[57.7%,100%]		[22.5%,63.7%]		[42.8%,83.3%]
	N=30		N=39		N=69	
High Effort	Incentivized Trust:	16.3%	Incentivized Trust:	5.9%	Incentivized Trust:	8.9%
		[8.1%,31.3%]		[3.5%,10%]		[5.9%,13.3%]
	Trustworthiness:	1.4%	Trustworthiness:	8.3%	Trustworthiness:	6.4%
		[0.6%,2.8%]		[4.9%,13.9%]		[4.2%,9.6%]
	Stated Trust:	7.3%	Stated Trust:	3.4%	Stated Trust:	4.9%
		[3.5%,14.7%]		[2.0%,5.7%]		[3.2%,7.4%]
	Risk:	44.4%	Risk:	26.1%	Risk:	32.4%
		[24.0%,74.9%]		[16.0%,41.0%]		[22.3%,45.8%]
	N=31		N=55		N=86	
Total	Incentivized Trust:	26.4%	Incentivized Trust:	10.3%	Incentivized Trust:	14.9%
		[16.7%,40.6%]		[7%,15.2%]		[11.1%,19.9%]
	Trustworthiness:	3.6%	Trustworthiness:	11.3%	Trustworthiness:	8.2%
		[2.2%,5.9%]		[7.6%,16.6%]		[6.1%,11.1%]
	Stated Trust:	5.4%	Stated Trust:	4.6%	Stated Trust:	4.9%
		[3.3%,8.9%]		[3.0%,6.8%]		[3.6%,6.7%]
	Risk:	68.9%	Risk:	32.1%	Risk:	45.1%
		[47.8%,93.6%]		[22.5%,44.8%]		[35.0%,57.1%]
	N=61		N=94		N=155	

Notes: The first row in each box corresponds to the error Ratio $\widehat{R}$ in incentivized trust estimated for the subjects in that box. The second row is the 95% confidence interval using a Fisher z transformation with the *corrci* command from Cox (2008). The following rows show the same statistics for measures of trustworthiness, stated trust and risk-aversion. The last row lists the number of observations in the box.

The first two rows of table 3 list the results for trust measured by the sender's behavior in the trust game.[16] The average amount of noise, about 15%, hides large variations across sub-groups. The noise figure, 37%, for the low effort/low CRT group is in particular non-negligible; in sharp contrast, the high effort/high CRT group is almost stable in test/retest. Effort and CRT are thus good predictors of the noise in the trust-game-based measure. For

[16]All confidence intervals in this subsection are estimated using a Fisher z transformation. Appendix E provides bootstrapped confidence intervals that are more conservative in our case.

trustworthiness, Table 3 reveals a moderate average level of noise. However, noise is almost non-existent for high effort/low CRT (under 2%) but is 15% for the high CRT/low effort sub-group. It is important to note that trustworthiness is elicited using the strategy method, i.e. the reported values are averaged across the potential amounts received. Furthermore, acting in the role of the receiver involves no uncertainty (strategic or otherwise): the outcome of the game is fully determined by the decision made by the receiver. We therefore expect CRT to have less influence than attention in the noise in this measure.

Measures of risk-attitude have been analyzed in detail, and the noise in these measures is typically found to be between 30% and 50%. We here provide corresponding figures to check that our subject pool displays common values, and to use these as a benchmark. Here too, average noise is not very representative of that in the sub-groups. For instance, in the extreme case of low CRT/low effort the risk measures are mostly noise. As such, compared to risk measures, trust measures are less prone to noise.

Contrary to the measures discussed above, stated trust is very stable, and varies only little by sub-group. Stated measures of trust are overall stable, as might be expected from the low average noise figure of 5%.

5. Discussion

5.1. Correcting errors in measurement without test/retest data

Instrumental methods can help correct errors in measurement with test/retest data. However, we do not always have such data at our disposal. Other methods allow to correct measurement errors without test/retest data but require having estimated (or having an idea of) the amount of noise. Our work gives an idea of the variance of measurement error in trust tasks and thus allows the use of such methods.

Plug-in estimator

A first methodology, that is relevant when having an explicit form of the estimator, is to plug the noise ratio (or the variance of the noise) estimator in the explicit estimator formula. For instance, when performing a simple linear regression of a dependent variable y on a noisy measure $x = x^* + \epsilon$ with $corr(x^*, \epsilon) = corr(y, \epsilon) = 0$, the OLS estimator $\hat{\beta}_{OLS}$ has an explicit formula (while being inconsistent).

$$\hat{\beta}_{OLS} \xrightarrow{n\to+\infty} \frac{cov(y,x)}{var(x)} = \frac{var(x^*)}{var(x^*)+var(\epsilon)}\,\beta = (1-R)\,\beta$$

Then a consistent estimator $\hat{\beta}_{CON}$ is simply obtained by plugging the estimator of the noise ratio in the formula :

$$\hat{\beta}_{CON} = \frac{1}{1-\hat{R}}\hat{\beta}_{OLS}$$

With a noise ratio of 15% (as found in subsection 4.2 for the amount sent in the trust game), you should multiply your OLS estimator (in simple linear regression) by about 1.18. If you believe your subject pool to be average in terms of cognitive skills and not too dedicated to the task, the noise ratio jumps up to 37% (see subsection 4.4). OLS coefficient should then be multiplied by about 1.59. Similarly, one can remove the bias in correlation estimation. If only one of both variables at stake is measured with error, you should multiply your correlation estimator by $\sqrt{1-\hat{R}}$. For multiple linear regressions, one could for instance use the `eivreg` or `sem` command from Stata software (see for instance Lockwood and McCaffrey, 2020).

Simulation Extrapolation (Simex) Method

A second method, that can be used without having explicit formulas, is the Simulation Extrapolation (Simex) Method. The simulation part consists in adding noise to the measure X at our disposal and estimating (for different levels of noise) the parameter of interests. The extrapolation part consists in extrapolating what would be an estimator of the parameter of interests without any noise (see Cook and Stefanski, 1994). For a practical implementation in Stata, you can for instance use the `simex` command introduced by Hardin et al. (2003). We invite economists willing to account for measurement errors in their research to check if their results hold when using the `eivreg`, `sem` or `simex` commands.

5.2. Measurement error and the "correlation puzzle"

Could the "correlation puzzle" (cf. section 2.2.) be explained by measurement error in elicited measures of trust? Noise in elicitation methods leads to under-estimate the correlation between two measures. Because different studies might be subject to a different extent to measurement errors, they do, however, not necessarily obtain a significant correlation.

Under the assumption of independence of errors across tasks, the ORIV methodology enables estimating the *true* correlations, i.e. without measurement error. We compare true correlations to the empirical correlations in our dataset. With two repeated observations for each variable, we can estimate four different correlations between two variables. We therefore computed the empirical correlations as the mean of all the pairwise correlations, and the "ORIV correlation".

We find that stated trust, trust behavior, and trustworthiness are all significantly and positively correlated (with coefficients between .38 and .56). As expected from the theory, the ORIV method pushes the correlation up by about 10%. In our experiment, correlation figures are already substantial before controlling for measurement errors. Furthermore, measurement error is quite low. So the ORIV method only produces limited changes.

In light of our results, the "correlation puzzle" remains. Indeed, some studies find no correlation. It would take massive amounts of noise to reconsider these results. For instance, assuming that both measures of trust include 50% of noise, the ORIV correlation would be twice as large as the uncorrected one. So very low correlation, say .02, would only move up to .04. The noise ratio found in the trust game (15%) could explain only a small attenuation bias in existing papers, and cannot account for the high variability of results in the literature. Thus, the investigation of the "correlation puzzle" must be continued.

We also obtained that risk-attitude is never significantly correlated with any measures of trust or trustworthiness. The absence of correlation remains even after correcting with the ORIV methodology. Hence, trust appears as a independent trait and not a by-product of risk-attitude.

6. Conclusion

Measures of trust are not all created equal regarding the amount of noise they produce. The trust-game measure entails more noise than the survey-based questions, but less than risk-attitude measures such as the Holt and Laury procedure. Furthermore, stated trust appears stable in test/retest, even for subgroups that are more prone to noise in the game. Our results thus shed light on the trade-off between incentives and noise for experimenters who wish to measure trust. If incentives are considered as important, the trust game can be used to elicit trust. However, even with highly-educated subjects the amount of noise is not negligible. On the contrary, for an experimenter who is worried about measurement error, surveys of interpersonal trust appear more appropriate.

A back-of-the-envelope calculation suggests that the general population would score lower than our subject pool.[17] The relevant measure of noise in the trust game, according to Table 3, would be in the range of 25 to 40%. In particular, "Lab in the field" experiments which target specific populations, sometimes with lower levels of education, and which occur in places with worse material conditions than university experimental laboratories, are likely to produce considerable noise affecting the elicited individual characteristics. Trust seems to be less problematic in this respect than risk attitudes, but the amounts of noise remain too large to be ignored.

Subjects' cognitive ability and effort seem to drive noise. Yet, our results do not explain the broad differences in noise between the trust game and the risk-aversion tasks. Finding the determinants of noise in incentivized measures would prove crucial to the design of future experiments and will be the subject of future research.

References

Aksoy, B., H. Harwell, A. Kovaliukaite, and C. Eckel (2018). Measuring trust: A reinvestigation. *Southern Economic Journal 84*(4), 992–1000.

Algan, Y. and P. Cahuc (2010). Inherited trust and growth. *American Economic Review 100*(5), 2060–92.

Algan, Y., P. Cahuc, and M. Sangnier (2016). Trust and the welfare state: the twin peaks curve. *Economic Journal 126*(593), 861–883.

Amador-Hidalgo, L., P. Brañas-Garza, A. M. Espín, T. García-Muñoz, and A. Hernández-Román (2021). Cognitive abilities and risk-taking: Errors, not preferences. *European Economic Review 134*, 103694.

Anderson, L. and J. Mellor (2009, 09). Are risk preferences stable? Comparing an experimental measure with a validated survey-based measure. *Journal of Risk and Uncertainty 39*, 137—160.

[17]The "low CRT" group provided on average 2.8 correct answers out of the six questions. In their meta-analysis Brañas-Garza et al. (2019) find an average score of 1.2 from three questions, which is somewhat lower.

Arrow, K. J. (1972). Gifts and exchanges. *Philosophy and Public Affairs 1*(4), 343–362.

Ashraf, N., I. Bohnet, and N. Piankov (2004). Is trust a bad investment? Working Paper, KSG.

Banerjee, R. (2018). On the interpretation of World Values Survey trust question-global expectations vs. local beliefs. *European Journal of Political Economy 55*, 491–510.

Bellemare, C. and S. Kröger (2007). On representative social capital. *European Economic Review 51*(1), 183–202.

Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, reciprocity, and social history. *Games and economic behavior 10*(1), 122–142.

Bishara, A. J. and J. B. Hittner (2017). Confidence intervals for correlations when data are not normal. *Behavior Research Methods 49*(1), 294–309.

Brañas-Garza, P., P. Kujal, and B. Lenkei (2019). Cognitive reflection test: Whom, how, when. *Journal of Behavioral and Experimental Economics 82*, 101455.

Capra, C. M., K. Lanier, and S. Meer (2008). Attitudinal and behavioral measures of trust: A new comparison. Working paper, Emory University.

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.

Cook, J. R. and L. A. Stefanski (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical association 89*(428), 1314–1328.

Cooper, D. and J. Kagel (2016, 01). Other regarding preferences: A selective survey of experimental results. *The handbook of experimental economics 2*, 217–289.

Cox, N. J. (2008). Speaking Stata: Correlation with confidence, or Fisher's z revisited. *The Stata Journal 8*(3), 413–439.

Falk, A., A. Becker, T. J. Dohmen, D. Huffman, and U. Sunde (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. Technical report, IZA Discussion Paper.

Fehr, E., U. Fischbacher, J. Schupp, B. Rosenbladt, and G. Wagner (2003). A nation-wide laboratory: Examining trust and trustworthiness by integrating behavioral experiments into representative surveys. *Schmoller's Jahrbuch 122*, 519–542.

Finucane, M. L. and C. M. Gullion (2010). Developing a tool for measuring the decision-making competence of older adults. *Psychology and Aging 25*(2), 271.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives 19*(4), 25–42.

Fukuyama, F. (1995). Social capital and the global economy. *Foreign Aff. 74*, 89.

Gillen, B., E. Snowberg, and L. Yariv (2019). Experimenting with measurement error: Techniques with applications to the Caltech cohort study. *Journal of Political Economy 127*(4), 1826–1863.

Glaeser, E. L., D. I. Laibson, J. A. Scheinkman, and C. L. Soutter (2000). Measuring trust. *Quarterly Journal of Economics 115*(3), 811–846.

Greiner, B. (2004). An online recruitment system for economic experiments. Technical report, MPRA.

Hardin, J. W., H. Schmiediche, and R. J. Carroll (2003). The simulation extrapolation method for fitting generalized linear models with additive measurement error. *The Stata Journal 3*(4), 373–385.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American Economic Review 92*(5), 1644–1655.

Horváth, R. (2013). Does trust promote growth? *Journal of Comparative Economics 41*(3), 777–788.

Johnson, N. D. and A. A. Mislin (2011). Trust games: A meta-analysis. *Journal of Economic Psychology 32*(5), 865–889.

Knack, S. and P. Keefer (1997). Does social capital have an economic payoff? a cross-country investigation. *The Quarterly journal of economics 112*(4), 1251–1288.

La Porta, R., F. Lopez-de Silanes, A. Shleifer, and R. W. Vishny (1997). Trust in large organizations. *American Economic Review 87*(2), 333.

Lazzarini, S., R. Madalozzo, R. Artes, and J. Siqueira (2003). Measuring trust: An experiment in Brazil. Technical report, IBMEC.

Lockwood, J. and D. F. McCaffrey (2020). Recommendations about estimating errors-in-variables regression in stata. *The Stata Journal 20*(1), 116–130.

Perez, F., G. Hollard, and R. Vranceanu (2021). How serious is the measurement-error problem in risk-aversion tasks? *Journal of Risk and Uncertainty 63*(3), 319–342.

Sangnier, M. (2013). Does trust favor macroeconomic stability? *Journal of Comparative Economics 41*(3), 653–668.

Sapienza, P., A. Toldra-Simats, and L. Zingales (2013). Understanding trust. *Economic Journal 123*(573), 1313–1332.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology 15*(1), 72–101.

Thöni, C., J.-R. Tyran, and E. Wengström (2012). Microfoundations of social capital. *Journal of Public Economics 96*(7-8), 635–643.

Toplak, M. E., R. F. West, and K. E. Stanovich (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & Reasoning 20*(2), 147–168.

Van Den Akker, O. R., M. A. Van Assen, M. Van Vugt, and J. M. Wicherts (2020). Sex differences in trust and trustworthiness: A meta-analysis of the trust game and the gift-exchange game. *Journal of Economic Psychology 81*, 102329.

Veszteg, R. F., Y. Funaki, and A. Tanaka (2015). The impact of the tohoku earthquake and tsunami on social capital in japan: Trust before and after the disaster. *International Political Science Review 36*(2), 119–138.

Yu, S., S. Beugelsdijk, and J. de Haan (2015). Trade, trust and the rule of law. *European Journal of Political Economy 37*, 102–115.

Zak, P. J. and S. Knack (2001). Trust and growth. *Economic Journal 111*(470), 295–321.

Appendix

Appendix A: Distributions of the main measures

This Appendix plots the distributions of our observations in the repeated measures for each of our main variables. Incentivized trust (Figure 2) follows a U-shaped distribution, with many people sending either none or all of their coins to the receiver, and few people sending amounts inbetween.

On the contrary, when answering the survey questions, our subjects are less concentrated on the extremes and appear more in the center of the distribution, producing a near normal distribution (Figure 1).

Concerning trustworthiness, the amounts sent back when participants play the role of the receiver are clustered around 0. There is also a peak at around one half, i.e. a fair redistribution of the gains (Figure 3).

Last, the Holt and Laury task produces a near normal distribution, as found in previous work (Figure 4).

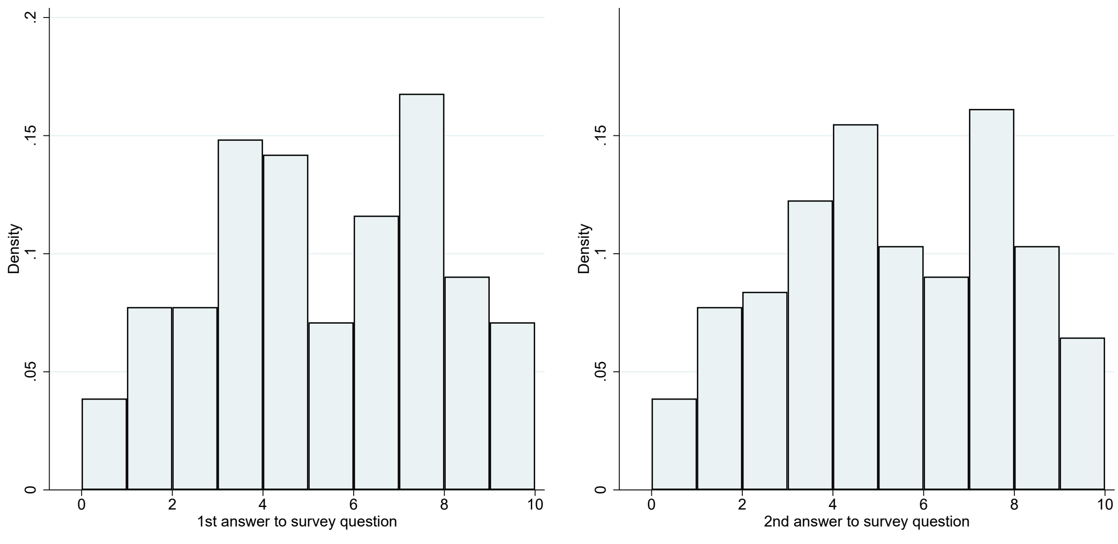


Fig. 1 The Distribution of Stated Trust in the Two Repetitions

Note: The histogram on the left (right) represents the density of the answers to the survey question the first (second) time it was asked.

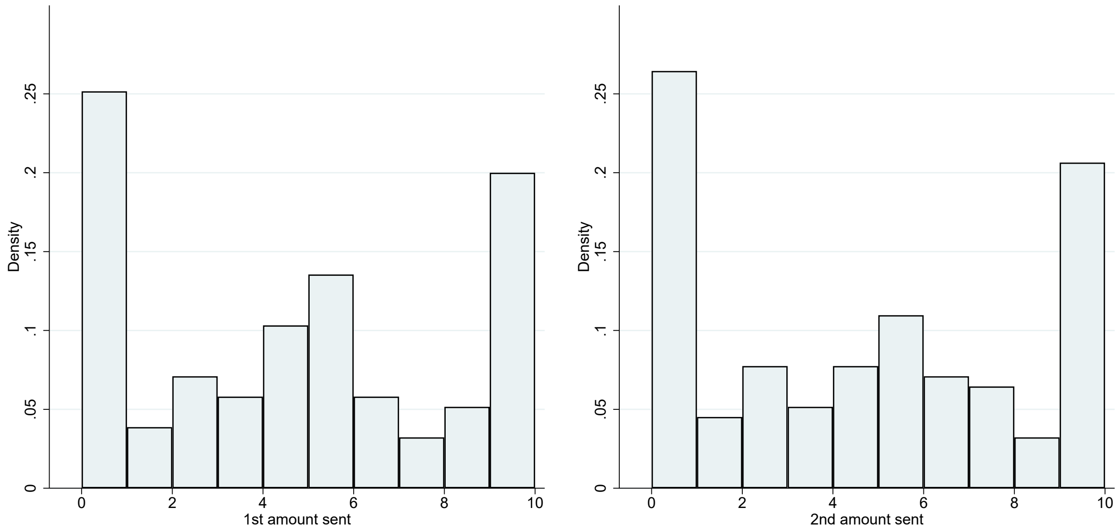


Fig. 2 The Distribution of the Amount Sent in the Two Repetitions

Note: The histogram on the left (right) represents the density of the amounts sent in the trust game the first (second) time it was played.

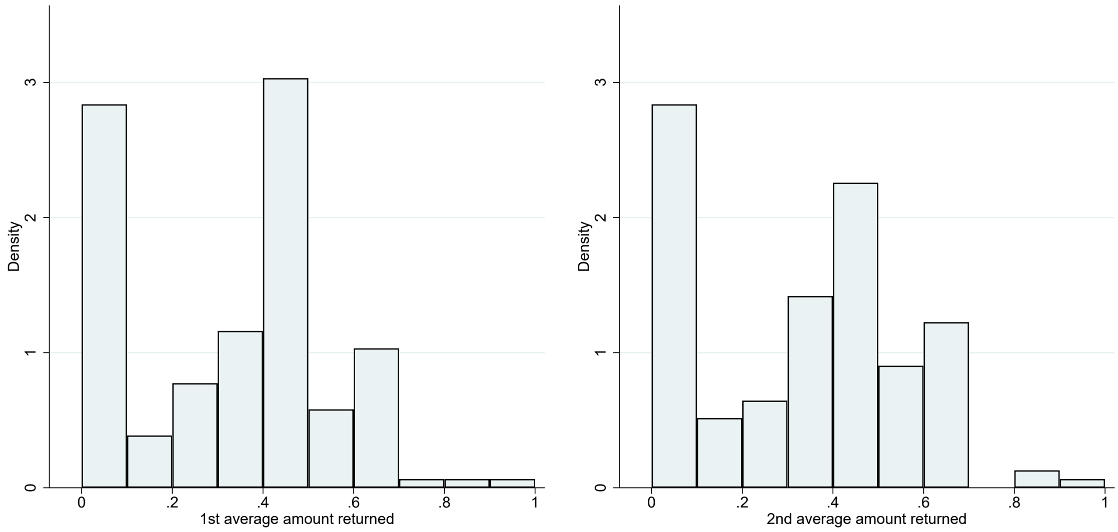


Fig. 3 The Distribution of the Average Amount Returned in the Two Repetitions

Note: The histogram on the left (right) represents the density of the average shares sent back in the trust game the first (second) time it was played.

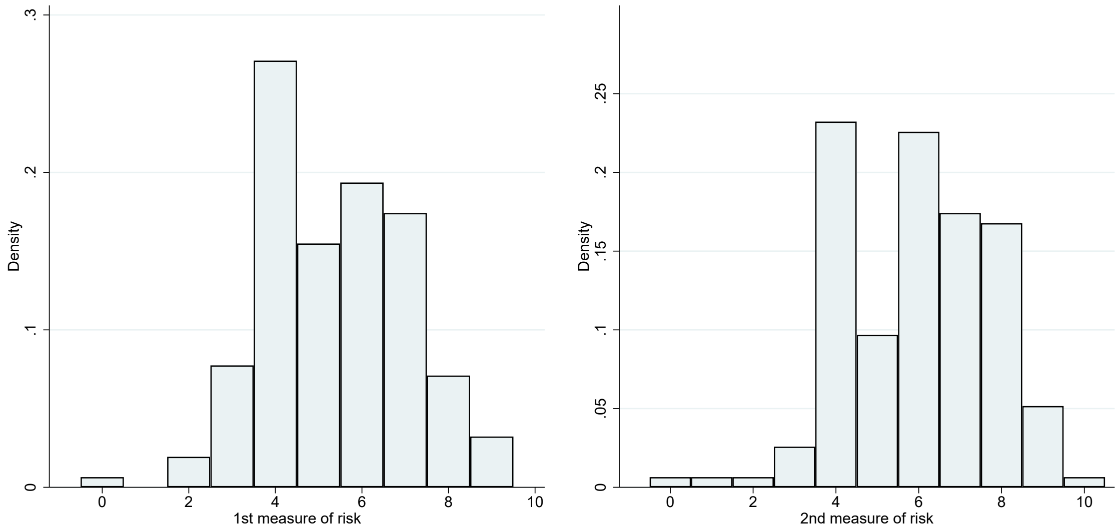


Fig. 4 The Distribution of Safe Choices in the Two Repetitions of the HL task

Note: The histogram on the left (right) represents the density of the number of safe choices in the first (second) HL task.

Appendix B: Repeated observations

This Appendix plots the individuals' repeated observations for our main variables. While the observations are close to the 45 degree line for stated trust (Figure 5) and trustworthiness (Figure 7), we observe much more noise for incentivized trust (Figure 6) and risk (Figure 8).

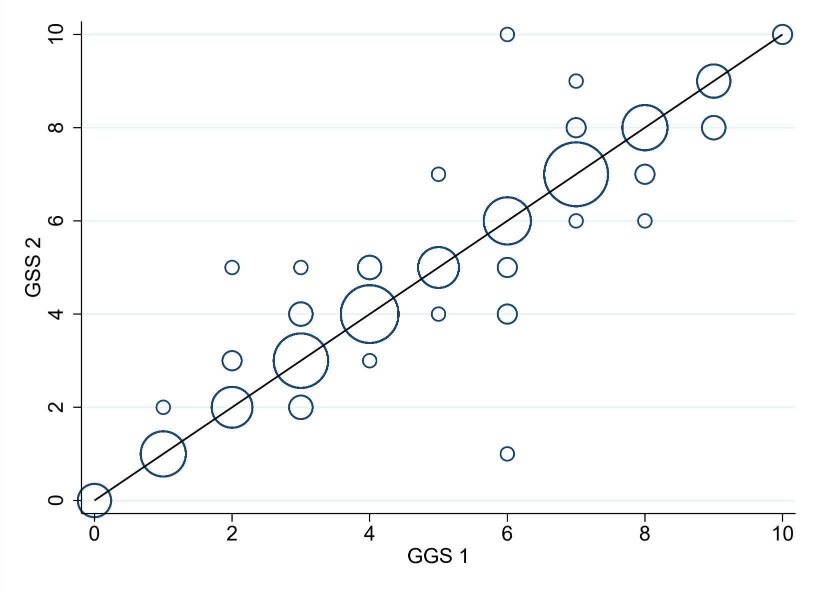


Fig. 5 Repetition of Stated Trust

Note: This plot represents the repeated answers to the survey question

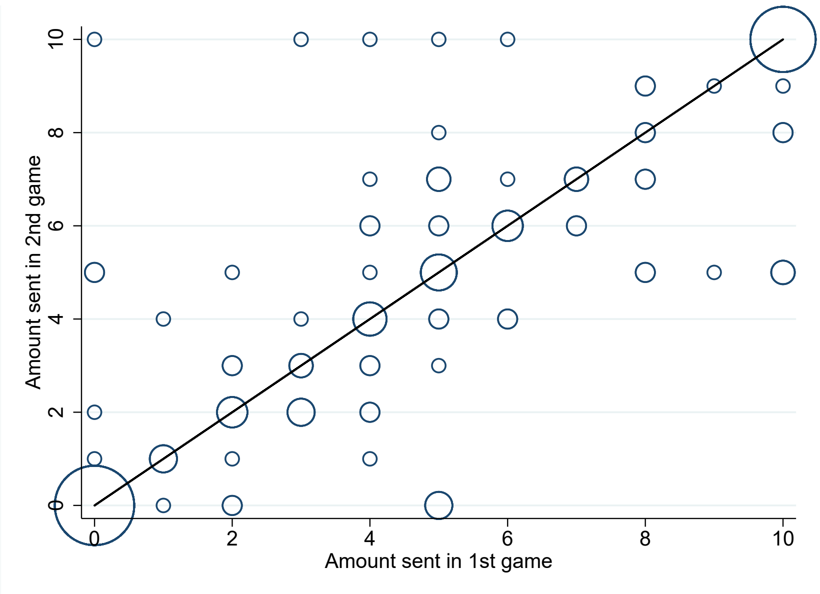


Fig. 6 Repetition of the Amount Sent

Note: This plot represents the repeated amounts sent in the trust game

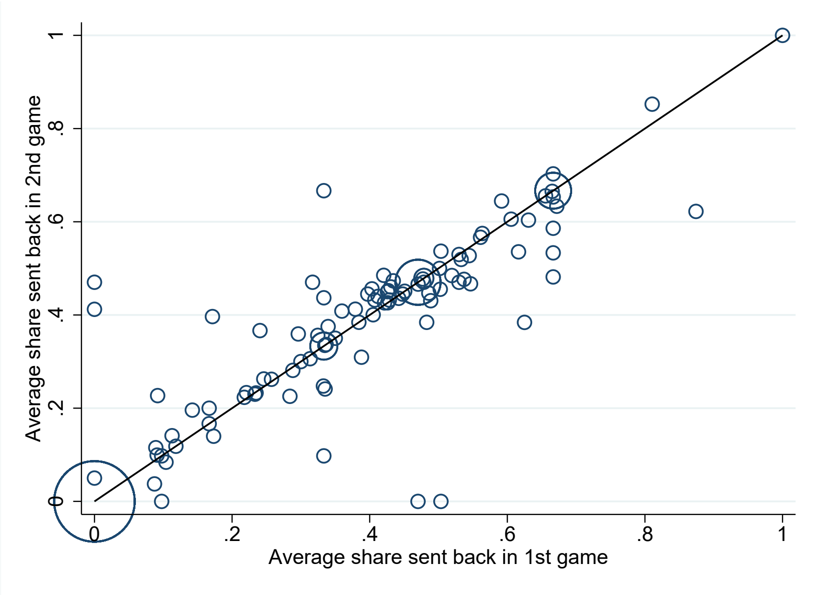


Fig. 7 Repetition of Trustworthiness

Note: This plot represents the repeated average shares sent back to the sender in the trust game.

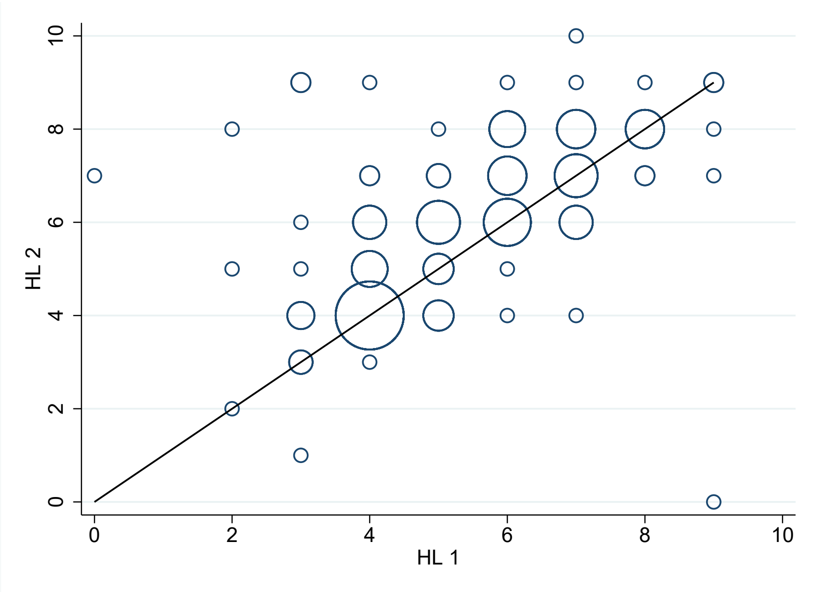


Fig. 8 Repetition of the HL Task

Note: This plot represents the repeated number of safe choices in the HL task

We use the risk-elicitation task from a previous experiment to estimate measurement error. This previous experiment took place just before the experiment that we analyze in this paper, such that there is about twenty minutes between the two risk-elicitation tasks. The incentives were the same in the two tasks. However, the units in question were different, with a different conversion rate between ECU and Euros. This may have nudged subjects to choose safer options in the second experiment. Nonetheless, our study focuses on the measurement error given by the random error and not the systematic error that could affect the first or second experiment. In other words, we can model what happens in both risk-elicitation tasks in the following way:

$$x_1 = \lfloor x^* + \epsilon_1 \rfloor \qquad \text{and} \qquad x_2 = \lfloor c + x^* + \epsilon_2 \rfloor$$

where c is the systematic error (a constant) that affects the second risk-elicitation task in comparison to the first one, and ϵ_1 and ϵ_2 are the two random errors. As we focus on random error, the fact that we have this systematic error is not a problem. Furthermore, parametric estimations that take this systematic error into account yield similar results to the non-parametric estimation we discuss in the paper (See Appendix D).

Appendix C: Trustworthiness

As discussed in the paper, the trustworthiness variable is the average proportion sent back by the receiver in the trust game. The ratio of measurement error observed in Table 2 is quite small, which can reflect that we consider the average of 10 figures. Constructing ten variables (Ret_x) corresponding to the trustworthiness for each choice in the strategy method (how many ECU the receiver sends back when the sender sends x ECU for $x \in [1; 10]$). Table 4 show the descriptive statistics for the 10 return variables.

Table 4: Amount returned in the ten cases in the strategy method

	First repetition	Second repetition
Ret_1	.75	.8
Ret_2	1.55	1.59
Ret_3	2.23	2.27
Ret_4	3.06	3.01
Ret_5	3.73	3.79
Ret_6	4.44	4.45
Ret_7	5.16	5.21
Ret_8	5.99	6.02
Ret_9	6.72	6.8
Ret_{10}	7.71	7.84

Table 5: Measurement error estimated for the different return variables

	$\widehat{\sigma_m}^2$	$\widehat{\sigma_\epsilon}^2$	$\widehat{R}$
Ret_1	.60	.10	.16
Ret_2	2.46	.29	.12
Ret_3	5.08	.46	.09
Ret_4	9.20	.93	.1
Ret_5	14.15	1.27	.09
Ret_6	19.75	1.76	.09
Ret_7	26.89	2.43	.09
Ret_8	36.08	3.36	.09
Ret_9	45.7	5.05	.11
Ret_{10}	60.45	7.42	.12

Appendix D: Parametric Estimation of measurement error

Using a similar method to that in Perez et al. (2021), we here provide the estimated variances and noise ratios using a parametric approach. We calculate these for stated trust, the amount sent in the Trust Game, and the Holt and Laury measure (allowing for a systematic error in HL). As noted above, as the return variable is the average of 10 figures, we have decided not to provide any parametric estimation of the measurement errors in this task. Parametric estimations lead to similar result, with a very slightly lower noise ratio being estimated in these three tasks.

Table 6: Measurement Error Parametric

	$\widehat{\sigma_m}^2$	$\widehat{\sigma_\epsilon}^2$	Noise Ratio $\widehat{R}$
Stated Trust	6.11	0.229	**3.60%**
Amount Sent	13.76	2.11	**13.3%**
HL	3.02	1.32	**43.9%**

Appendix E: Sub-group Analysis with Bootstrapped Confidence Intervals

There is an ongoing debate about the method that should be used to calculate confidence intervals for empirical correlations (see the recent paper by Bishara and Hittner, 2017). The Fisher z transformation to obtain confidence intervals produces not very conservative figures, in particular when the variables are not Gaussian. Here our variables are bounded, so the confidence intervals may not be totally inappropriate. We nonetheless provide the sub-group analysis with bootstrapped confidence intervals that are less conservative (and thus larger, especially with small n).

Table 7: Estimation of the Noise Ratio in Trust Behavior by Sub-groups

	Low CRT	High CRT	Total
Low ScoreGrid	37.4% [3.1%,71.7%] N=30	17.9% [6.8%,29.1%] N=39	24.3% [9.8%,38.8%] N=69
High ScoreGrid	16.3% [2.8%,29.8%] N=31	5.9% [0.1%,11.8%] N=55	8.9% [3%,14.8%] N=86
Total	26.4% [8.1%,44.8%] N=61	10.3% [4.8%,15.9%] N=94	14.9% [8.2%,21.7%] N=155

Notes: The first row in each box corresponds to the error Ratio $\widehat{R}$ estimated on the subjects in that box. The second row is the 95% Bootstrap confidence interval, and the third row the number of observations.

Table 8: Estimation of the Noise Ratio in Stated Trust by Sub-groups

	Low CRT	High CRT	Total
Low ScoreGrid	3.7% [0.3%,7.0%] N=30	6.3% [0.0%,13.3%] N=39	5.1% [1.1%,9.1%] N=69
High ScoreGrid	7.3% [0.0%,19.0%] N=31	3.4% [0.0%,6.8%] N=55	4.9% [0.0%,9.9%] N=86
Total	5.4% [0.0%,11.2%] N=61	4.6% [1.1%,8.0%] N=94	4.9% [1.7%,8.1%] N=155

Notes: The first row in each box corresponds to the error Ratio $\widehat{R}$ estimated on the subjects in that box. The second row is the 95% Bootstrap confidence interval, and the third row the number of observations.

Table 9: Estimation of the Noise Ratio in the Risk-Attitude Measure by Sub-groups

	Low CRT	High CRT	Total
Low ScoreGrid	92.6% [26.4%,100%] N=30	39.2% [13.3%,65%] N=39	61.2% [25.2%,97.2%] N=69
High ScoreGrid	44.4% [15.0%,73.9%] N=31	26.1% [6.9%,45.3%] N=55	32.4% [16.2%,48.5%] N=86
Total	68.9% [29.0%,100%] N=61	32.1% [16.3%,48.0%] N=94	45.1% [25.8%,64.3%] N=155

Notes: The first row in each box corresponds to the error Ratio $\widehat{R}$ estimated on the subjects in that box. The second row is the 95% Bootstrap confidence interval, and the third row the number of observations.

Table 10: Estimation of the Noise Ratio in Trustworthiness by Sub-groups

	Low CRT	High CRT	Total
Low ScoreGrid	5.3%	15.2%	10.2%
	[0.8%,9.8%]	[1.4%,29%]	[2.7%,17.7%]
	N=30	N=39	N=69
High ScoreGrid	1.4%	8.3%	6.4%
	[0.2%,2.5%]	[0.0%,18.0%]	[0.0%,13.3%]
	N=31	N=55	N=86
Total	3.6%	11.3%	8.2%
	[1.1%,6.1%]	[2.8%,19.8%]	[3.0%,13.5%]
	N=61	N=94	N=155

Notes: The first row in each box corresponds to the error Ratio $\widehat{R}$ estimated on the subjects in that box. The second row is the 95% Bootstrap confidence interval, and the third row the number of observations.

Appendix F: Screenshots from the Experiment

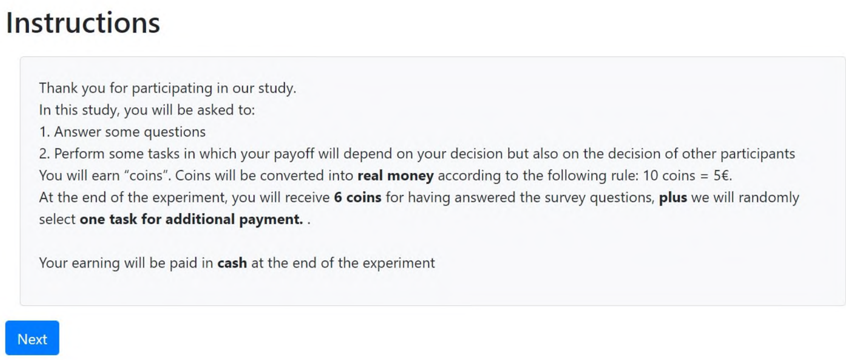


Fig. 9 Instructions for All Experiments

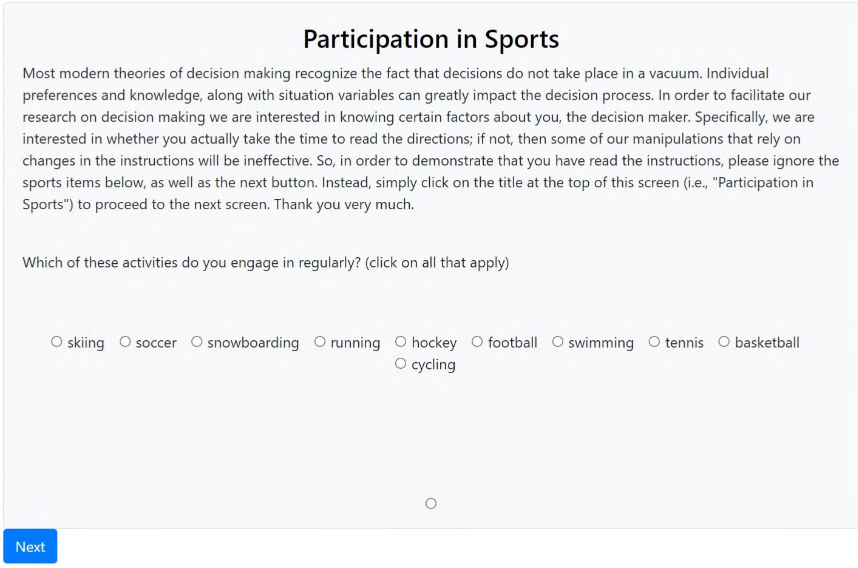


Fig. 10 Question to Focus Participants' Attention

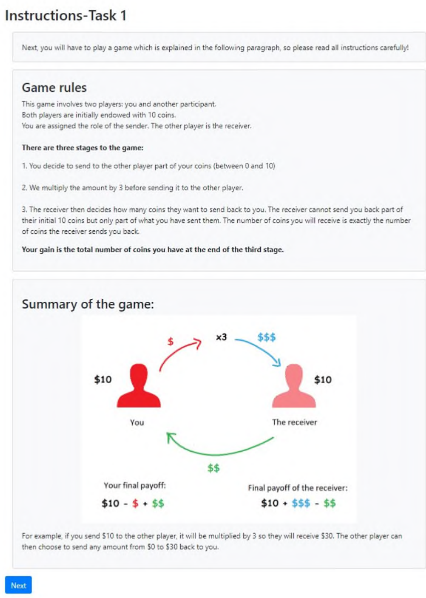
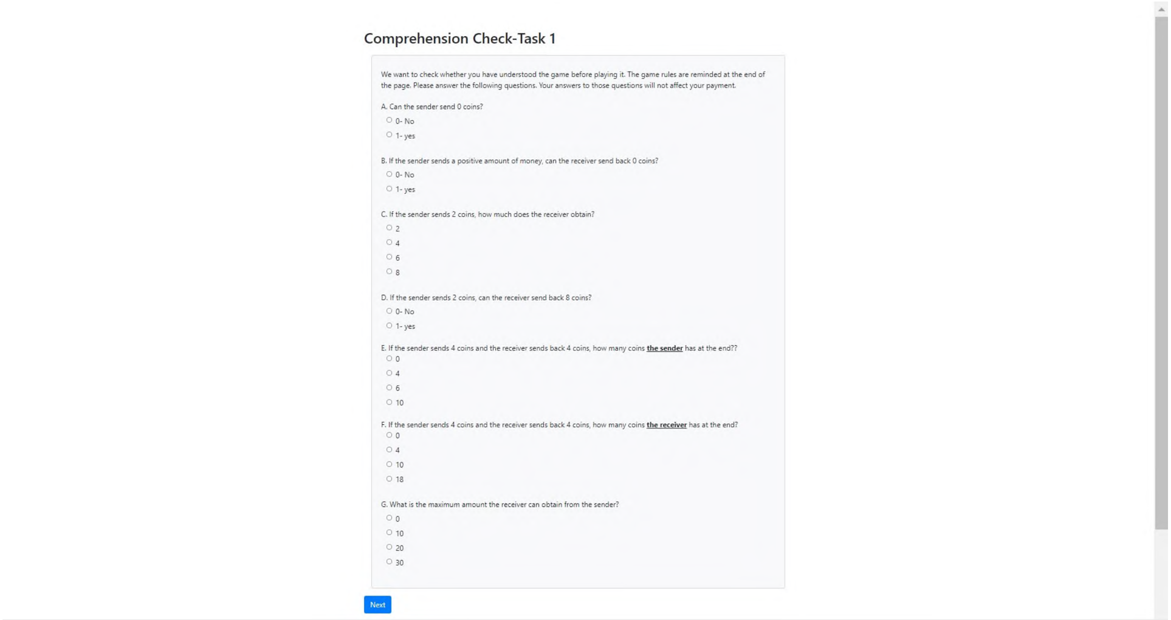


Fig. 11 Trust Game Instructions



Comprehension Check-Task 1

We want to check whether you have understood the game before playing it. The game rules are reminded at the end of the page. Please answer the following questions. Your answers to those questions will not affect your payment.

A. Can the sender send 0 coins?
- 0- No
- 1- yes

B. If the sender sends a positive amount of money, can the receiver send back 0 coins?
- 0- No
- 1- yes

C. If the sender sends 2 coins, how much does the receiver obtain?
- 2
- 4
- 6
- 8

D. If the sender sends 2 coins, can the receiver send back 8 coins?
- 0- No
- 1- yes

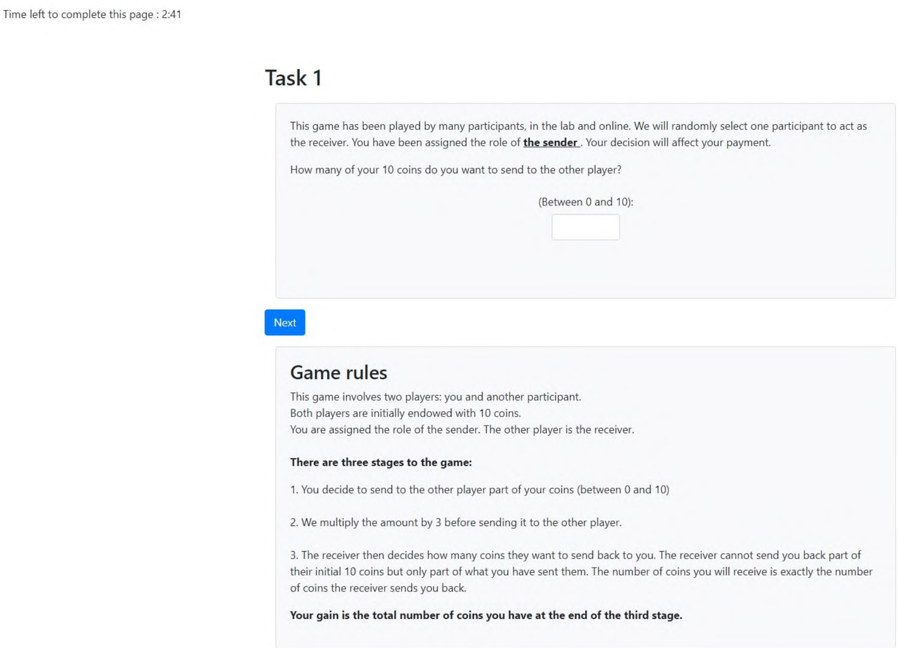
E. If the sender sends 4 coins and the receiver sends back 4 coins, how many coins **the sender** has at the end??
- 0
- 4
- 6
- 10

F. If the sender sends 4 coins and the receiver sends back 4 coins, how many coins **the receiver** has at the end?
- 0
- 4
- 10
- 18

G. What is the maximum amount the receiver can obtain from the sender?
- 0
- 10
- 20
- 30

Next

Fig. 12 Trust Game Comprehension Check



Time left to complete this page : 2:41

Task 1

This game has been played by many participants, in the lab and online. We will randomly select one participant to act as the receiver. You have been assigned the role of **the sender** . Your decision will affect your payment.

How many of your 10 coins do you want to send to the other player?

(Between 0 and 10):

Next

Game rules

This game involves two players: you and another participant.
Both players are initially endowed with 10 coins.
You are assigned the role of the sender. The other player is the receiver.

There are three stages to the game:

1. You decide to send to the other player part of your coins (between 0 and 10)

2. We multiply the amount by 3 before sending it to the other player.

3. The receiver then decides how many coins they want to send back to you. The receiver cannot send you back part of their initial 10 coins but only part of what you have sent them. The number of coins you will receive is exactly the number of coins the receiver sends you back.

Your gain is the total number of coins you have at the end of the third stage.

Fig. 13 Answer to Trust Game as Sender

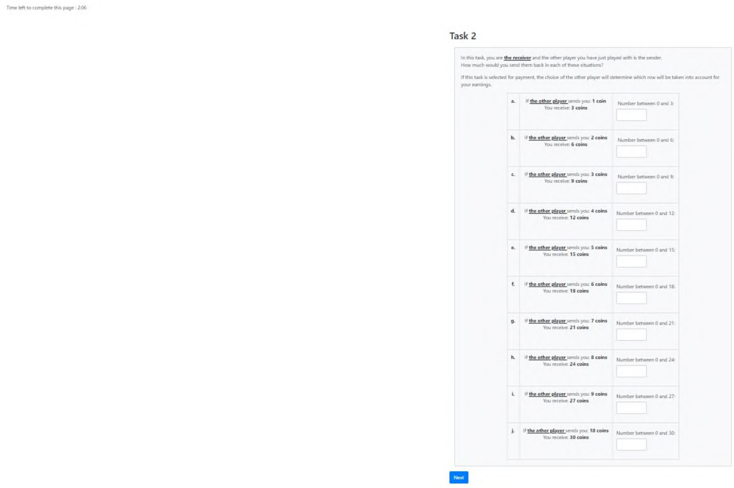
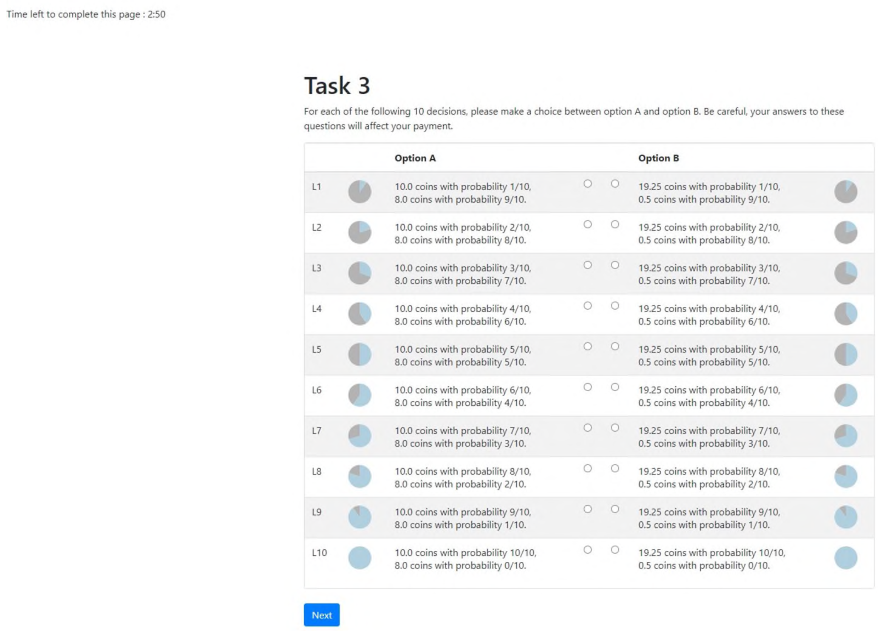


Fig. 14 Answer to Trust Game as Receiver

Fig. 15 Instructions for the Risk-Preference Elicitation Task



Time left to complete this page : 2:50

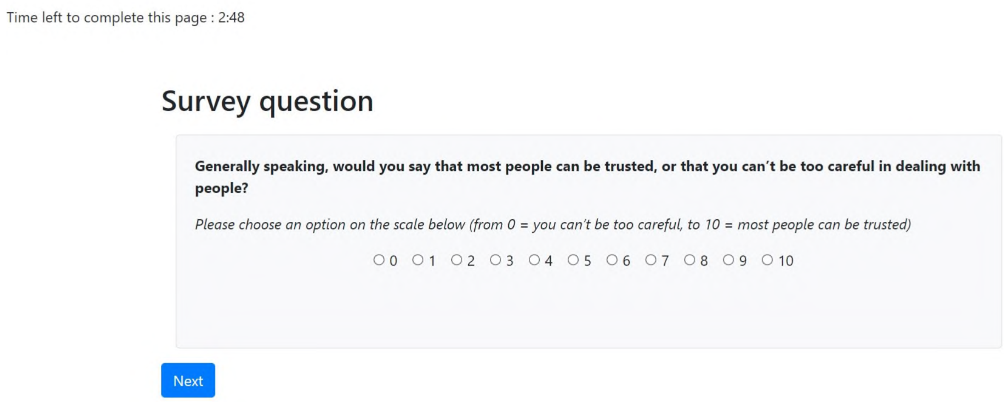
Task 3

For each of the following 10 decisions, please make a choice between option A and option B. Be careful, your answers to these questions will affect your payment.

	Option A	Option B
L1	10.0 coins with probability 1/10, 8.0 coins with probability 9/10.	19.25 coins with probability 1/10, 0.5 coins with probability 9/10.
L2	10.0 coins with probability 2/10, 8.0 coins with probability 8/10.	19.25 coins with probability 2/10, 0.5 coins with probability 8/10.
L3	10.0 coins with probability 3/10, 8.0 coins with probability 7/10.	19.25 coins with probability 3/10, 0.5 coins with probability 7/10.
L4	10.0 coins with probability 4/10, 8.0 coins with probability 6/10.	19.25 coins with probability 4/10, 0.5 coins with probability 6/10.
L5	10.0 coins with probability 5/10, 8.0 coins with probability 5/10.	19.25 coins with probability 5/10, 0.5 coins with probability 5/10.
L6	10.0 coins with probability 6/10, 8.0 coins with probability 4/10.	19.25 coins with probability 6/10, 0.5 coins with probability 4/10.
L7	10.0 coins with probability 7/10, 8.0 coins with probability 3/10.	19.25 coins with probability 7/10, 0.5 coins with probability 3/10.
L8	10.0 coins with probability 8/10, 8.0 coins with probability 2/10.	19.25 coins with probability 8/10, 0.5 coins with probability 2/10.
L9	10.0 coins with probability 9/10, 8.0 coins with probability 1/10.	19.25 coins with probability 9/10, 0.5 coins with probability 1/10.
L10	10.0 coins with probability 10/10, 8.0 coins with probability 0/10.	19.25 coins with probability 10/10, 0.5 coins with probability 0/10.

Next

Fig. 16 Answer to the Risk-Elicitation Task



Time left to complete this page : 2:48

Survey question

Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?

Please choose an option on the scale below (from 0 = you can't be too careful, to 10 = most people can be trusted)

○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

Next

Fig. 17 Stated Trust Question

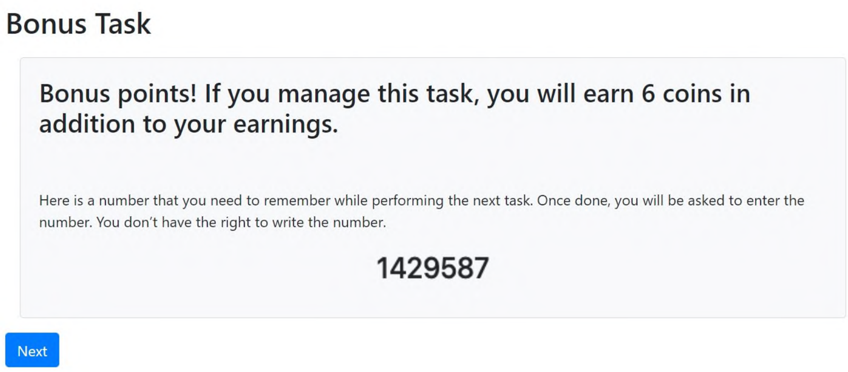


Fig. 18 Distraction Task, Number Task

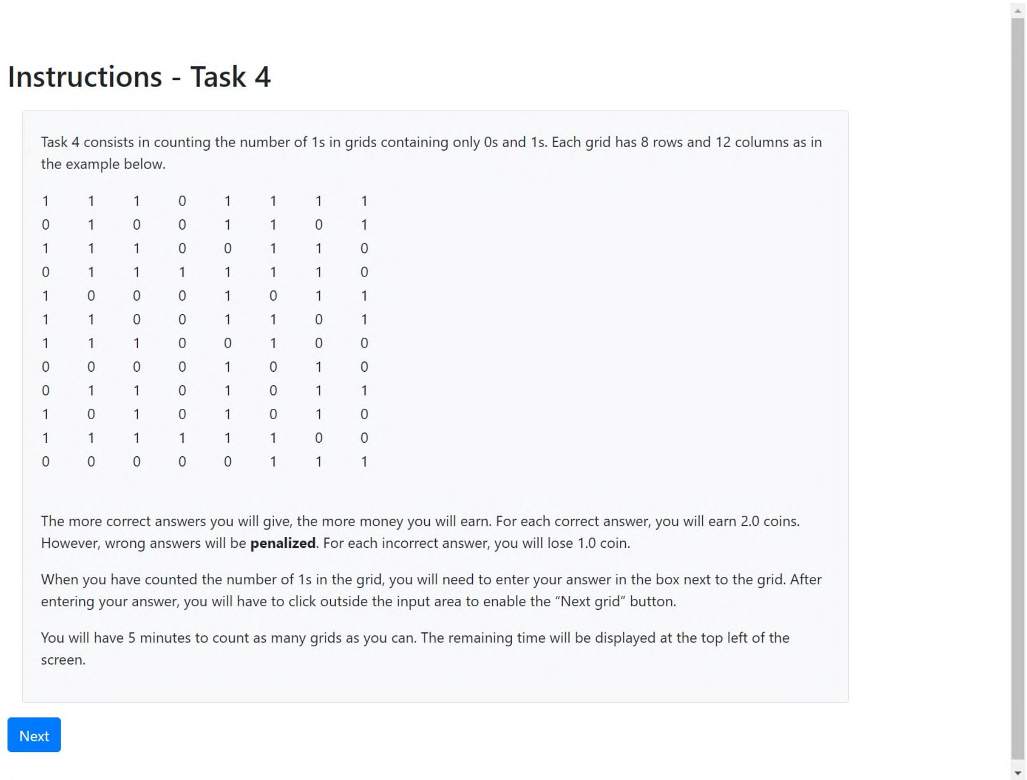


Fig. 19 Distraction Task, Matrix Instructions

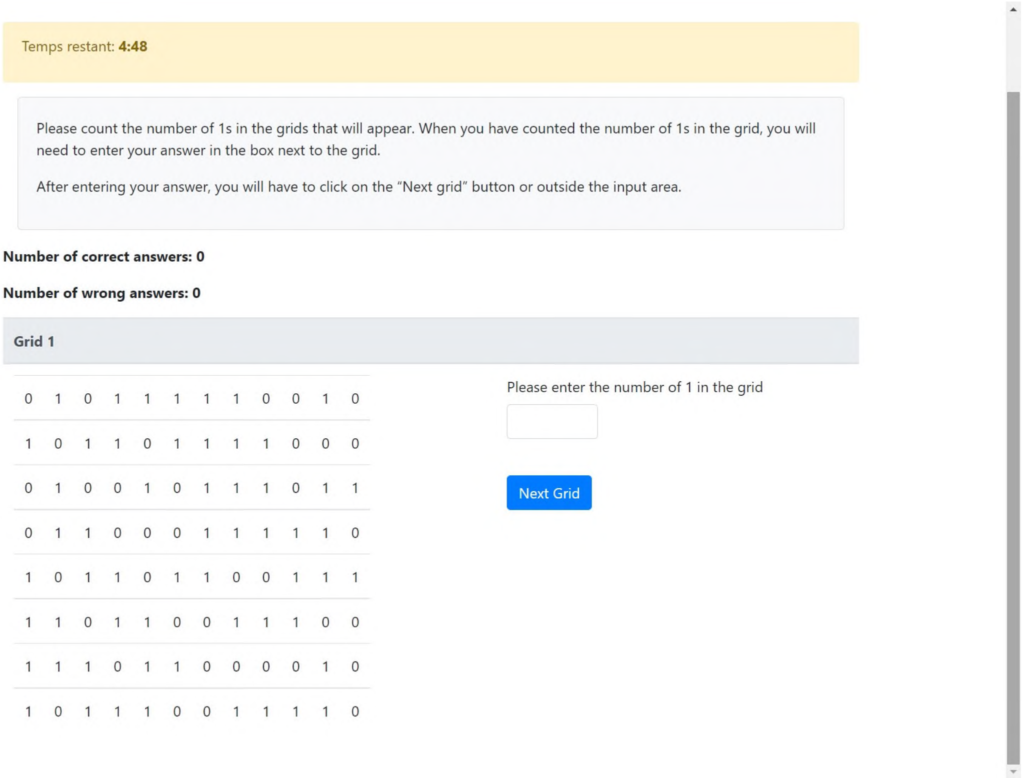


Fig. 20 Distraction Task, Matrix

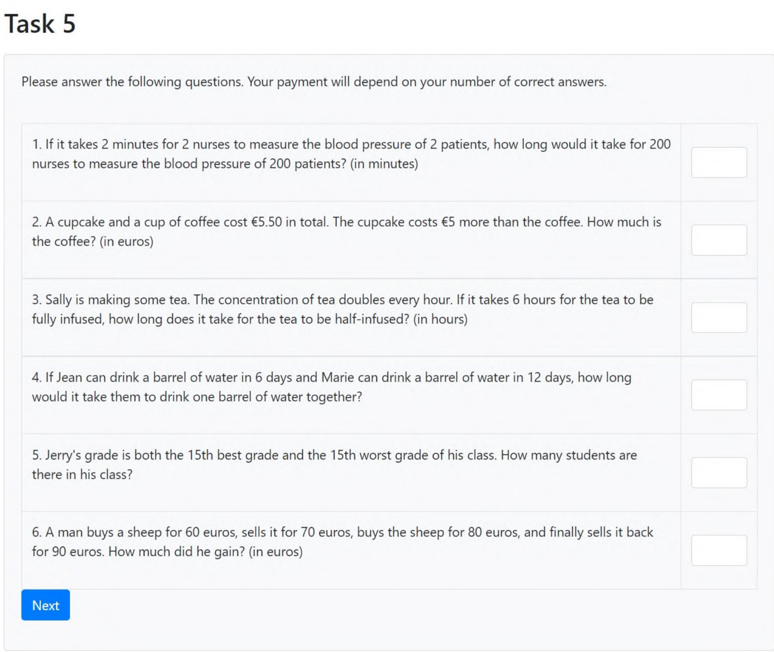
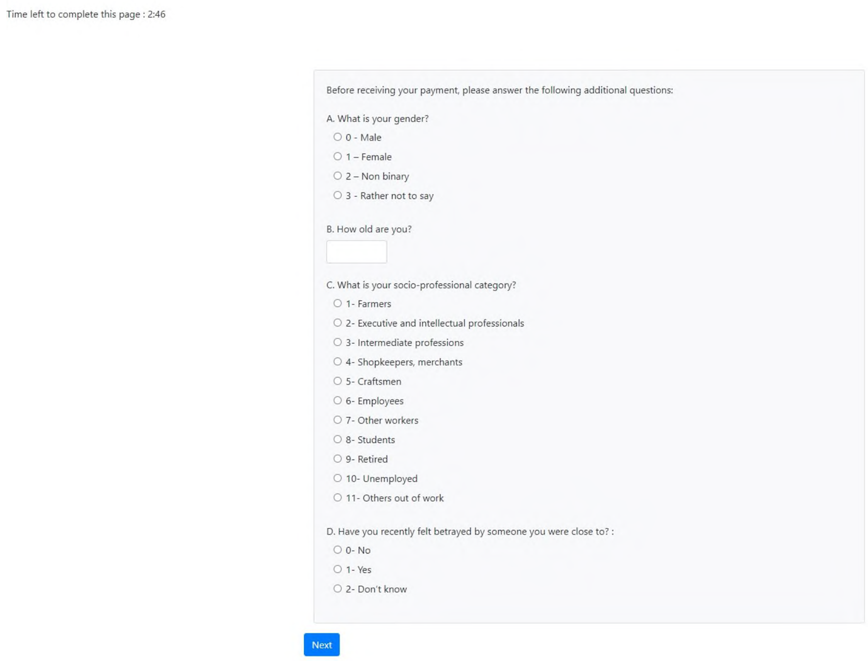


Fig. 21 CRT Questions



Time left to complete this page : 2:46

Before receiving your payment, please answer the following additional questions:

A. What is your gender?

- 0 - Male
- 1 – Female
- 2 – Non binary
- 3 - Rather not to say

B. How old are you?

C. What is your socio-professional category?

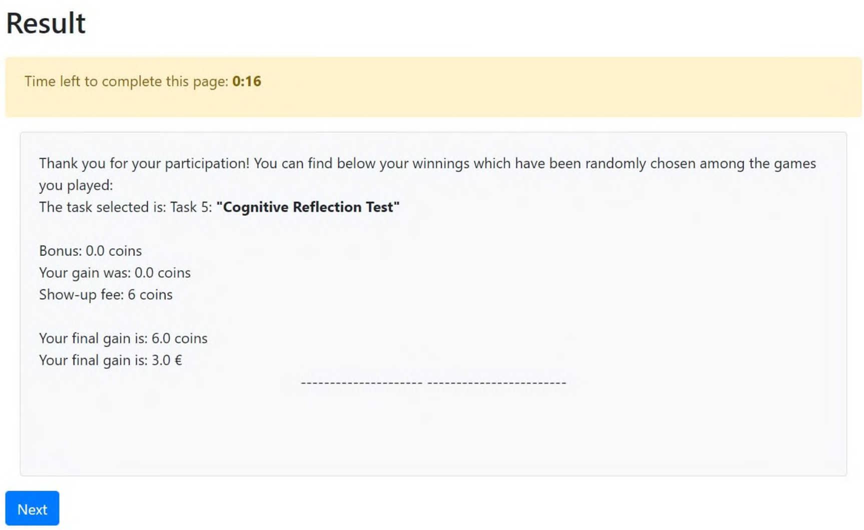
- 1- Farmers
- 2- Executive and intellectual professionals
- 3- Intermediate professions
- 4- Shopkeepers, merchants
- 5- Craftsmen
- 6- Employees
- 7- Other workers
- 8- Students
- 9- Retired
- 10- Unemployed
- 11- Others out of work

D. Have you recently felt betrayed by someone you were close to? :

- 0- No
- 1- Yes
- 2- Don't know

Next

Fig. 22 Socio-Economic Questions



Result

Time left to complete this page: **0:16**

Thank you for your participation! You can find below your winnings which have been randomly chosen among the games you played:
The task selected is: Task 5: **"Cognitive Reflection Test"**

Bonus: 0.0 coins
Your gain was: 0.0 coins
Show-up fee: 6 coins

Your final gain is: 6.0 coins
Your final gain is: 3.0 €

-------------------- -----------------------

Next

Fig. 23 Individual Result Presentation

Appendix G: Correlated errors in measurement

To estimate the amount of noise entailed in behavioral measures, we assumed that errors are not correlated. In this subsection, we show that if subjects exhibit "experiment" or "day" fixed effects, what we would estimate with our method is a lower bound of the variance of noise in measurement.

Indeed, assume that the noise is of the form: $\epsilon_{ik} = \eta_i + \tilde{\epsilon}_{ik}$ with a noise fixed effect η_i that is not correlated with a random noise $\tilde{\epsilon}_{ik}$.

Assume also that we observe test/retest data

$$X_{i1} = X_i^* + \eta_i + \tilde{\epsilon}_{i1} \quad \text{and} \quad X_{i2} = X_i^* + \eta_i + \tilde{\epsilon}_{i2}$$

We estimate the variance of the noise by studying the difference between the two measures, such that:

$$Var(X_{i1} - X_{i2}) = Var(\tilde{\epsilon_1} - \tilde{\epsilon_2}) = 2Var(\tilde{\epsilon}) < 2Var(\epsilon)$$

If the errors were correlated, the variance of the noise we estimated would therefore correspond to the variance of $\tilde{\epsilon}$ and not the variance of the total error ϵ. Our measurement error ratio would be under-estimated. Questioning the hypothesis of uncorrelated errors only supports the claim that we obtained a lower bound noise ratio.

CRESTCENTER FOR RESEARCH IN ECONOMICS AND STATISTICS

CREST
Center for Research in Economics and Statistics
UMR 9194

5 Avenue Henry Le Chatelier
TSA 96642
91764 Palaiseau Cedex
FRANCE

Phone: +33 (0)1 70 26 67 00
Email: info@crest.science
https://crest.science/

The Center for Research in Economics and Statistics (CREST) is a leading French scientific institution for advanced research on quantitative methods applied to the social sciences.

CREST is a joint interdisciplinary unit of research and faculty members of CNRS, ENSAE Paris, ENSAI and the Economics Department of Ecole Polytechnique. Its activities are located physically in the ENSAE Paris building on the Palaiseau campus of Institut Polytechnique de Paris and secondarily on the Ker-Lann campus of ENSAI Rennes.