

Bellemare, Charles; Kröger, Sabine

**Working Paper**

**On Representative Trust**

SFB 373 Discussion Paper, No. 2003,27

**Provided in Cooperation with:**

Collaborative Research Center 373: Quantification and Simulation of Economic Processes,  
Humboldt University Berlin

*Suggested Citation:* Bellemare, Charles; Kröger, Sabine (2003) : On Representative Trust, SFB 373 Discussion Paper, No. 2003,27, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin,  
<https://nbn-resolving.de/urn:nbn:de:kobv:11-10050278>

This Version is available at:

<https://hdl.handle.net/10419/22242>

**Standard-Nutzungsbedingungen:**

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

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

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

**Terms of use:**

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

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

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

# On Representative Trust\*

Charles Bellemare<sup>†</sup>      Sabine Kröger<sup>‡</sup>

July 16, 2003

## Abstract

This paper combines an economic experiment with survey data to investigate determinants of trust and trustworthiness in the Dutch society. We contrast the inferences which can be made on the trust propensity using stated and revealed measures and we test for participation bias in our experiment. We find that middle aged and educated individuals trust relatively more but are relatively less trustworthy. The effect of age and religion on trust is shown to depend heavily on whether experimental or survey trust measures are used. We find no evidence of participation bias in any experimental decisions.

JEL Codes: Z13, C90, C10

Keywords: Measuring trust and trustworthiness, Experiments, Representative samples.

---

\*Financial support from CentER and CentERdata is gratefully acknowledged. We thank Hanneke Dam, Marcel Das and Corrie Vis of CentERdata for their support throughout the experiment. We are grateful to Oliver Kirchkamp, Arthur van Soest, Eric van Damme, Jan Magnus, Pierre-Carl Michaud, Wieland Müller and Karim Sadrieh for helpful comments. Useful comments were made by participants at the NAKE Research Day in Amsterdam, the ENDEAR Workshop in Jena, the ESA European Meeting in Strasbourg, the Max-Planck-Institute in Jena, the CIRANO in Montréal, the DIW in Berlin, the ENTER Jamboree 2003 meeting and the Social Capital Workshop in Tilburg, the Spring Meeting of Young Economists in Leuven, and the Economics seminar in Maastricht. The second author gratefully acknowledges financial support by the Deutsche Forschungsgemeinschaft through Sonderforschungsbereich 373, Humboldt Universität zu Berlin, and the EU-TMR ENDEAR Network grant (FMRX-CT98-0238). The usual disclaimer applies.

<sup>†</sup>Tilburg University, CentER, The Netherlands,  
{email: c.bellemare@uvt.nl}.

<sup>‡</sup>Humboldt Universität zu Berlin, Germany and Tilburg University, CentER, Netherlands,  
{email: skroeger@wiwi.hu-berlin.de}

# 1 Introduction

It is increasingly argued that a nation's social capital can influence its economic performance. Although there is an ongoing debate over what constitutes social capital (Bowles and Gintis, 2002; Durlauf 2002), there seems to be a consensus that both average societal trust and trustworthiness are two important components. The transactions cost paradigm remains the traditional way of thinking about the mechanism by which both these components of social capital affect economic performance. When societal trust and trustworthiness levels are high, transactions costs are low which makes organizations and governments more efficient which ultimately leads to better economic performance.

The research on social capital started with the influential work of Putnam (1993) who found a strong correlation between measures of civic engagement and government quality across regions in Italy. The association of social capital with growth started with the work of Knack and Keefer (1997), and Zak and Knack (2001) who find that a one-standard deviation increase in the World-Value Survey (WVS) trust index increases economic growth by more than one-half of a standard deviation. La Porta, de Silanes, Shleifer, and Vishny (1997) find that across countries, a one-standard deviation increase in the same measure of trust increases judicial efficiency by 0.7 of a standard deviation and reduces government corruption by 0.3 of a standard deviation. These empirical facts rest on the WVS trust index constructed by drawing in each participating country a random sample of participants who are asked to answer, amongst others, the following question

**WVS trust question** Generally speaking would you say that most people can be trusted or that you cannot be too careful in dealing with people?

- 1.) Most people can be trusted.
- 2.) You have to be very careful.
- 3.) I do not know.

The WVS reports for each country the percentage of responders who indicated that "Most people can be trusted".

Because of the strong correlation between measures of social capital and economic growth, and since age and income distributions of many Western societies are predicted to evolve over time in very alarming ways (Gruber and Wise, 2001; Gottschalk and Smeeding, 1997), it becomes relevant for policy makers to investigate how average trust and trustworthiness in their population are shifted as a result of changes in age, education, income, and past life experiences of individuals. To perform these measurements, two essential features must be combined: 1) to have a random sample drawn from a country's population, and 2) being able to measure trust and trustworthiness with little error. The motivation for the first condition follows from the law of large numbers, whereas sample average trust and trustworthiness are consistent estimates of their population counterparts. The second requirement follows from the fact that the estimated parameters we make inferences on will, in general, be biased if trust and trustworthiness are measured with error (Bound, Brown, and Mathiowetz, 2001).

The empirical literature has up to now been unable to meet both requirements simultaneously. The empirical methods employed so far to analyze individual trust are survey questionnaires and economic experiments. Alesina and La Ferrara (2002) use answers to the WVS trust question to investigate determinants of trust in the United States. The main advantage of their approach is it allows to make population inferences by observing the behavior of a randomly drawn sample of individuals from that population. The main drawback is that researchers run the risk of collecting answers to a vague and hypothetical question which can create a discrepancy between someone's answers and his actual behavior. Thus, part of the variation in responses may be attributed to differences in interpreting who compromises "most people", differences in what is meant to trust someone etc. Moreover, variations in response may also arise because individuals do not answer truthfully to the question.

Economic experiments have the virtue of countering the effects associated with survey data by observing the actual behavior of individuals placed in a context which is under experimental control. These experiments offer an attractive alternative to surveys, given one can design an experiment which captures the essential features of trust and trustworthiness which are of interest. The seminal experiment of Berg, Dickhaut, and McCabe (1995) (hereafter BDMc) remains today the main experimental

design used to test for the presence of trust and trustworthiness (more details on the game are given later). The general results of the BDMc game are that people place trust in others, but it is ambiguous whether or not this trust pays. These results have since been shown to be robust to several framing effects (Ortmann, Fitzgerald, and Boeing, 2000) and role reversals (Burks, Carpenter, and Verhoogen, 2003). The main drawback of these types of experiments is that subjects are generally drawn from homogenous pools of university students. These subjects lack the required variation in background characteristics to measure how these variables influence trust and trustworthiness.

The analysis of determinants of trustworthiness is even less documented than trust and relies mostly on analyzing responders' behavior in trust games such as the BDMc game. Hence, this form of analysis suffers from the same drawbacks outlined above. Our little knowledge of the determinants of trustworthiness at the population level is source of concern as recent research has argued that trustworthiness could be "the" economically relevant component of social capital to understand the process of economic development (Francois and Zabochnik, 2002).

This paper makes three important contributions. First, we combine the strengths of survey and experimental methods in a straightforward way by having a random sample of the Dutch population play a computerized version of the BDMc trust game. This allows us to touch on several related issues. First and foremost, we add to the scarce body of knowledge on determinants of trust and trustworthiness by estimating age, education, and other life experience effects from experimental trust responses. The key results are that the age and education level of subjects influence trust and trustworthiness in very different ways. Specifically, we find that there is an inverted U shape relation between trust and education, and trust and age, while both relationships are U shaped with respect to trustworthiness. The later finding contrasts with some of the existing relationships found in the social capital literature. The second advantage of combining survey and experimental methods is that we are able to test the assumption of parallelism between the lab and the field. Up till now, this assumption has generally been tested with newspaper experiments (e.g., see Bosch-Domènech, Montalvo, Nagel, and Satorra (2002) for a survey of newspaper experiments) and internet experiments (e.g., Lucking-Reiley, 1999). Because newspaper readers or internet users are not

generally representative of a nation's population, these mediums prevent population inferences which are the primary concern of our paper.<sup>1</sup>

The second contribution of the paper builds on the seminal work of Glaeser, Laibson, Scheinkman, and Soutter (2000) (hereafter GLSS) who, among other things, invalidate the use of survey trust measures on the basis that they do not predict well trust responses in the lab. We show that the predictive power of survey trust is strongly related to the sample used and the experimental design. The immediate consequence of this is that experimenters can influence the predictive power of survey trust by appropriately choosing subjects and experimental designs. This leads us to conclude that the predictive power of survey trust questions is insufficient to validate or invalidate their use. We propose a new approach to compare both trust measures, which consists of testing whether survey and experimental approaches are equally informative on the determinants of trust. Our evidence shows that the effect of some background characteristics, especially religion and education, can change dramatically when using stated rather than experimental trust responses.

Finally, by giving our subjects the choice to participate and by observing the characteristics of those who refuse to do so, we are in the unique position to test for participation bias in our experiment. If participants have for example above average taste for gambling and risk, or higher cognitive abilities than non-participants, and these unobservable attributes are correlated with the outcomes that are measured in the lab, a participation bias will be present. Because experiments typically do not observe non-participants, tests of participation bias in experiments are very limited. Eckel and Grossman (2000) find some evidence on the presence of participation bias in a classroom experiment by comparing responses of student volunteers and "pseudo"-volunteers. Their results are specific to the game they use (a dictator game) and hold

---

<sup>1</sup>Three noteworthy experiments have recently been run with representative samples. Harrison, Lau, and Williams (2002) use a random sample of the Danish population to investigate the heterogeneity in individual discount rates. Hey (2002) used the CentERpanel of Tilburg University (more on this panel later on) to have a random sample of the Dutch population play an experiment on decision making under risk and uncertainty. Fehr, Fischbacher, Rosenblatt, Schupp, and Wagner (2002) report about a "preliminary analysis" of a "first implementation" of an interview based trust game with the German Socio-Economic Panel.

only for student populations. In this paper we provide the first full fledged test of participation bias in experimental economics. We do not find any evidence suggesting that the trust and trustworthiness behavior of participants in our experiment differ in any way from that of randomly selected subjects.

The remainder of the paper is organized as follows. Section 2 describes the design of the experiment, the experimental procedure, and our sample. Section 3 discusses the empirical results on trust. Findings on trustworthiness are presented in section 4. Section 5 reports results for participation bias while section 6 discusses and concludes.

## 2 The experimental design and the sample

The recruitment of our subjects was made by CentERdata, the survey research institute of Tilburg University in the Netherlands. The main activity of CentERdata is to manage and carry out panel surveys through a telepanel: the CentERpanel (hereafter CP), consisting of approximately 2000 representative Dutch households. Every Friday, CP's household members receive a questionnaire which they are asked to fill in at any time between Friday and Tuesday of the following week. This questionnaire is filled at home either on a computer or on a television set which is connected to a set-up box linking the household to the CentERdata server. In order to keep the sample representative of the Dutch population, low income households without a computer or a television set are given the necessary equipment in order to complete the weekly questionnaire.<sup>2</sup>

There are many reasons why the CP is an attractive medium to conduct experiments. First, it gives us access to a representative sample of a population, which is the key feature of this study. Second, because participants answer questions on a computer or a television set, we are able to replicate as closely as possible the environment of a laboratory experiment, which simplifies comparisons of our results with those of the

---

<sup>2</sup>For a description of the recruitment, sampling methods, and past usages of the CentERpanel see: [www.centerdata.nl](http://www.centerdata.nl). Children below 16 years of age as well as immigrants are excluded from the panel. The latter for the reason being that their language proficiency in Dutch makes it difficult for them to answer the questions on a weekly basis.

existing literature. Third, because participants communicate with CentERdata, the experiment is double blind as participants were told that they will be anonymously matched and that their identities would not be revealed to the experimenters. Finally, as CentERdata reimburses the weekly telephone costs for answering the questionnaire by crediting CentERpoints to their private bank accounts four times a year, our participants are already familiar to payment in fictitious currency.<sup>3</sup> This allows us to use CentERpoints as the experimental currency unit and reimburse our participants in a very convenient way.

Our design closely follows BDMc.<sup>4</sup> A sender<sup>5</sup>  $S$  and a responder  $R$  were both endowed with 500 CentERpoints.  $S$  could send money to  $R$  from his endowment. We discretized the choice set of  $S$  to 11 amounts  $y^E \in \{0, 50, \dots, 450, 500\}$ . The amount  $S$  sent was doubled by the experimenters and added to  $R$ 's endowment. We measured responses using the strategy method by which  $R$  was asked to return an amount to  $S$ , contingent on each of the 11 possible amounts he might receive from  $S$ . The response which corresponded to the actual decision of  $S$  was chosen to be the effective action and determined the payoff of both participants. After all participants made their decisions,  $S$  and  $R$  players were randomly matched and payoffs were computed based on the decisions of the pair. The final payoffs were computed as follows:  $S$  received the initial 500 CentERpoints reduced by the amount sent  $y^E$  plus the amount received from  $R$ , while  $R$  received his initial endowment of 500 CentERpoints, the amount sent by  $S$  multiplied by 2 minus the amount returned to  $S$ .

The strategy method was chosen to overcome the difficulty of having CP members play together in real time. This method has several additional advantages. First, it facilitates data acquisition as the complete strategy plan for all 11 possible amounts received is elicited. Second, as our game may seem complex to some subjects, the strategy method requires that people thoroughly familiarize themselves with the ram-

---

<sup>3</sup>The exchange rate is 1 CentERpoint = 0.01 Euro.

<sup>4</sup>Computer screens of the original experiment (in Dutch) are available upon request. The translated text of all screens are enclosed in appendix B.

<sup>5</sup>For ease of reading we keep the terms "sender" and "responder" for the different roles. In the experiment we omitted suggestive labels and referred to the person itself or to his opponent as "the matched panel member."



ifications of all choices, so that we do not retrieve data from uninformed subjects.<sup>6</sup>

Under the assumption that both players maximize their monetary payoffs, the Nash equilibrium of the game is for  $S$  to send nothing to  $R$ , as  $R$ 's dominant strategy is to return nothing to  $S$ . Hence, observing increasing positive amounts sent is interpreted as evidence that people increasingly trust others. Likewise, observing increasing amounts returned is taken as evidence of the existence of increasing trustworthiness. It is important to stress that repeated game effects, retaliation strategies, and game experience effects are deliberately excluded by our experimental design. Thus, one can think of the current design as measuring the basic trust propensity of an individual at a given point in time.<sup>7</sup>

$S$  and  $R$  were additionally asked to state their beliefs about their partners' action. These questions were asked after both players made their decisions in order to circumvent the possibility that belief elicitation induces non-cooperative behavior when asked before the play of the game (Croson, 2000). We elicit senders' beliefs with two questions. The first question asks to state how much they think  $R$  will return to them. The second question asks them to state what they think the average  $S$  player will send. The latter question is intended to capture behavior directed towards some social norm. Responders on the other hand simply had to state how much units they thought of receiving from senders. This concluded the experimental part of the session.

All players were then asked to answer two survey questions. The first question asked players to state their average experience with trust

**Lifetime trust experience question** In the past, when you trusted someone, was your trust usually rewarded or usually exploited?

(Always rewarded) 1 , 2 , 3 , 4 , 5 , 6 , 7 (Always exploited).

---

<sup>6</sup>There is weak evidence suggesting that a hot environment triggers stronger responses in two player games. Brandts and Charness (2000) find that the strategy method and the hot environment do not yield significant different responses in two simple sequential two player games.

<sup>7</sup>Results from repeated trust experiments can be found in Anderhub, Engelmann, and G  th (1999) and Willinger, Keser, Lohmann, and Usunier (2003).

This question will be used to test for the presence of state dependent behavior whereby differences in past experiences with trust may lead to different experimental decisions.

The second question was the WVS trust question presented in the introduction. This question will allow us to compare the inferences on trust which can be made using stated and revealed decisions. Contrary to GLSS, our subjects answered the WVS trust question after having made their experimental decisions. This has the benefit of not framing the experimental decision as one involving trust. The disadvantage is that answers to the WVS trust question may be influenced by the experimental decisions which were made before, which would complicate the comparisons of our experimental and survey trust measures. In the next section, we develop a simple econometric model which allows to test for the presence of misreported answers to the WVS trust question, and examines to which extent misreporting was influenced by the experimental decision. As will be shown, we do not find evidence suggesting that our sequence of tasks influenced answers to the WVS question.

Two weeks after the experiment, each participant received feedback information on the outcome of the game and their final payoff which was later credited to their CentER bank accounts. The experiment was conducted in two sessions, in the 31st and the 36th weeks of the calendar year 2002. Individuals contacted had to read an opening screen informing them that they were selected to participate in an experiment conducted jointly by a team of university researchers. A detailed description of the game followed with the mode of payments. Each person was informed that conditional on their participation, they would be randomly matched to one of the roles. The role was revealed to panel members once they had agreed to participate. We contacted 541 panel members from which 42 declined to participate.<sup>8</sup> Of the 499 panel members who completed the experiment, 276 were *S* players and 223 were *R* players.

Table 1 gives the description of the variables and descriptive statistics of the 541 household members contacted for senders, receivers and non-participants. The means

---

<sup>8</sup>14 non-participants initially accepted to play but eventually backed out of the experiment after having observed the roles they were assigned to play. It is interesting to note that 11 out of those 14 panel members were assigned to the role of responders.

of most variables are relatively identical across non-participants, senders, and responders. 63.7% of the persons contacted were heads of households and most players either had a secondary or vocational training degree. Catholics and protestants are the two most important religious communities in the sample and their relative weights in the three participation categories are very similar. Two notable differences across the three groups concern work propensity and age. Non-participants are on average 10 years older than both senders and responders. This age effect is also reflected in a higher labor market retirement frequency and lower labor work participation.

Despite all the advantages of using the CP, one limitation is that panel members were not constrained to complete the experiment in a limited amount of time, giving them the opportunity to seek advice in order to make more informed choices. As a result, a high game time would be an indication of collective decision making. CentERdata keeps track of the time taken to complete the questionnaires from the time subjects log in the CentERdata network. Table 1 also displays some statistics on completion times for both types of players and non-participants. As expected, non-participants have the lowest participation time in the experiment, with a median time slightly greater than a minute. The median time taken by  $R$  players is greater than that by  $S$  players, a result primarily due to the fact that  $R$  players made their decisions using the time-intensive strategy method. Since the majority of subjects took less than 10 minutes to complete the experiment, it seems unlikely that collective decision making is present in the data.

### 3 Results on Trust

The distribution of amounts sent in the experiment is shown in figure 1. The two distinctive features of this distribution are 1) the majority of subjects send positive amounts 2) the distribution is heavily skewed to the left, with a mode at 5, the equal split category. The shape of this distribution is familiar to that usually found in lab-experiments with student samples (BDMc; Ortmann, Fitzgerald, and Boeing, 2000) but differs greatly from that of GLSS, which was heavily skewed to the right, with most subjects sending the maximal possible amount. We will try to reconcile the

differences between GLSS and our data below.

We assume that senders have a continuous unobserved latent trust propensity  $T_i^*$ . This propensity is heterogeneously distributed in the population according to

$$T_i^* = \mathbf{x}_i^t \beta + \varepsilon_i^t \quad (1)$$

where  $\mathbf{x}_i^t$  is a vector of observed characteristics of sender  $i$ ,  $\beta$  is a vector of unknown slope parameters, and  $\varepsilon_i$  is a random term capturing unobserved heterogeneity across individuals. Our experimental trust measure  $y_i^E$  is ordinal and discrete. The ordered probit model is adequate to analyze this type of data (Maddala, 1983). However, the ordered probit model requires a sufficient amount of observations in each discrete category to estimate nuisance threshold parameters. As can be seen from figure 1, categories 300 to 450 CentERpoints have very little observations. In our empirical application, we merge these categories and estimate an ordered probit model with eight categories.

The first 2 specifications of table 2 present ordered probit regressions. The first specification uses as regressors a standard set of background characteristics supplemented by reported life experience with trust (TRUSTEXP), subjects' beliefs about the amount they expect to be returned to them (STHINK), and the average amount they expect other senders will send (SMEANS). The second specification omits beliefs.<sup>9</sup>

In both specifications, we do not find gender of subjects to influence trust. This contrasts with the earlier findings in the social capital literature which showed that women are less involved in organizations (Glaeser, Laibson, and Sacerdote, 2002) and that women trust less than men (Alesina and La Ferrara, 2000). We also find that family size, whether an individual is retired from the labor force or is not working do not correlate with trust.

The estimated age effect is robust and of similar magnitude across both specifications. Both age parameters are significant. Based on specification 1 and holding other factors constant, trust is seen to increase until the age of 30, beyond which it

---

<sup>9</sup>We have experimented with a specification including cross-terms but none was found to be statistically significant.

starts to decline. This reconfirms the inverted-U shape pattern usually found in the social capital literature (Putnam, 2000; Glaeser, Laibson, and Sacerdote, 2002) although those studies report that social capital reaches a high at 45 years of age. Education also has an inverted-U profile. We find that individuals with secondary and technical training are more likely to trust than subjects with either low education levels (the omitted category) and subjects with university degrees. Comparisons of specification 1 with specification 2 show that the education relationship is also robust to inclusions of beliefs. Glaeser, Laibson, and Sacerdote (2002) find a positive relation between education and organizational membership but do not report results which would indicate the presence of non-linearities. Subjects' religious beliefs were classified in three categories, protestants, catholics, and atheists. We find no evidence that either catholics or protestants trust differently than atheists (the omitted category), which squares with results found by Alesina and La Ferrara (2002).

Both belief variables, STHINK and SMEANS, have positive effects on trust and are highly significant. These results indicate that senders who expected to receive more sent more, and senders who thought other senders would send more increased their amount sent. The former result captures expectations of the subjects.<sup>10</sup> The latter result can be interpreted as a social norm as individuals tend to partly emulate what they expect others to do. In order to assess the contribution of beliefs to the empirical model, we computed a likelihood-ratio test comparing specifications 1 and 2. The test value of 232.12 (5%  $\chi^2$  critical value of 5.99) indicates that apart from being statistically significant, beliefs substantially improve the predictive fit of the model.

Finally, we asked subjects to report their lifetime experience with trust. This was done to investigate whether a form of state dependence existed, whereas an individual who is dissatisfied with his experiences with trust is less likely to trust in the experiment.<sup>11</sup> Our results show that the state of trust individuals perceive themselves to be

---

<sup>10</sup>It has been argued that the causality may be in the opposite direction (Selten and Ockenfels, 1998, p. 526-529). We do not investigate these issues in this paper.

<sup>11</sup>This definition of state dependence differs from that used in labor economics. Labor economists are generally interested in whether or not individuals in a state of unemployment are more likely to be unemployed in the future. In our experiment, we do not investigate whether having trusted in the past increases the likelihood of trusting in the experiment.

in does not significantly correlate with experimental trust.

### **3.1 Comparing experimental and survey trust measures**

In this section, we sequentially address the following questions: 1) Do answers to the WVS trust question predict well experimental trust? 2) Do survey and experimental trust measures provide the same information on the determinants of trust? It is important to stress that both questions are complementary. The first question is relevant if we are interested in predicting experimental trust responses. However, it leaves open the reasons of what drives the possible correlation between both variables. The second question asks whether we can extract the same information from both set of measures.

#### **3.1.1 Prediction**

The question of whether or not answers to the WVS trust question predict experimental trust has been addressed by GLSS. Running a linear regression of experimental trust of answers to the WVS trust question and a set of covariates, they find that answers to the WVS do not significantly explain their experimental trust outcomes. The main explanations given for this result are that the WVS question is vague, hypothetical, and likely to be misreported. To investigate which of these explanations is more relevant, we first follow GLSS by adding answers to the WVS trust question (WVS) as a regressor in our experimental trust equation. This corresponds to specification 3 of table 2. Contrary to GLSS, we find that answers to the WVS trust question do significantly explain experimental trust. Furthermore, all other parameter estimates of the model are robust to the inclusion of this variable.

How can we rationalize the finding that answers to the WVS trust question predicts well our experimental trust and not that of GLSS? In our view, the two elements explaining the differences between both studies are the amount of background information on subjects which is available, and differences in experimental designs. The role of each of these factors can be better understood by analyzing the following system of

linear equations<sup>12</sup>

$$y_i^E = \mathbf{x}_i^t \beta^E + \xi y_i^{WVS} + \varepsilon_i^E \quad (2)$$

$$y_i^{WVS} = \mathbf{x}_i^t \beta^{WVS} + \varepsilon_i^{WVS} \quad (3)$$

Equation (2) is the linear version of equation (1), where  $y_i^{WVS}$  denotes answers to the WVS question,  $\varepsilon_i^E$  captures the unobserved determinants of experimental trust and  $\xi$  measures the predictive power of answers to the survey trust question. The second equation relates answers to the WVS trust question to  $\mathbf{x}_i^t$  and an unobserved component  $\varepsilon_i^{WVS}$ . It is important to point out that the result which we present below will hold in general for any other measure or linear combination of survey trust and social capital measures which are used as predictors. Substituting (3) in (2) and rearranging yields

$$y_i^E = \mathbf{x}_i^t \beta^{E+} + \xi \varepsilon_i^{WVS} + \varepsilon_i^E \quad (4)$$

where  $\beta^{E+} = \beta^E + \xi \beta^{WVS}$ . From equation (4) we see that the value of  $\xi$  is identified from the variation between  $y_i^E$  and  $\varepsilon_i^{WVS}$ . Partialling out the covariates  $\mathbf{x}_i^t$ , it is straightforward to show that for the linear model, the probability limit of the estimated predictive effect  $\hat{\xi}$  is

$$p \lim \hat{\xi} = \xi + \left( \frac{\mathbf{V}(\varepsilon_i^E | \mathbf{x}_i^t)}{\mathbf{V}(\varepsilon_i^{WVS} | \mathbf{x}_i^t)} \right)^{1/2} \mathbf{Cor}(\varepsilon_i^E, \varepsilon_i^{WVS} | \mathbf{x}_i^t) \quad (5)$$

where  $V(\varepsilon_i^E | \mathbf{x}_i^t)$  and  $V(\varepsilon_i^{WVS} | \mathbf{x}_i^t)$  are the variances of  $\varepsilon_i^E$  and  $\varepsilon_i^{WVS}$  conditional on  $\mathbf{x}_i^t$  and  $Cor(\varepsilon_i^E, \varepsilon_i^{WVS} | \mathbf{x}_i^t)$  denotes the correlation between both unobserved components for a given set of covariates.

The impact of the amount of background information available and the experimental design on the estimated predictive power of survey trust measures emerge from equation (5). First, the amount of information on the characteristics of subjects which is available to experimenters will play an important, as any omitted characteristics remotely correlated with survey and experimental trust will be captured by  $\varepsilon_i^E$

---

<sup>12</sup>In the probit model,  $E(y | \mathbf{X} = \mathbf{x}) = \Phi(\mathbf{x}'\beta)$  is approximately linear in  $\mathbf{x}'\beta$  for individuals answering 1 to the WVS trust question with probability between 20% and 80% (e.g., Ruud, 2000). In general, the average probability of answering 1 to the WVS questions lies between 45% and 55% which implies that the linear model holds for most individuals in the sample.

and  $\varepsilon_i^{WVS}$ . The higher the number of common unobserved characteristics, the higher  $Cor(\varepsilon_i^E, \varepsilon_i^{WVS} | \mathbf{x}_i^t)$  will be. It is important to note that a stronger correlation will amplify the differences in estimates of  $\hat{\xi}$  between studies which have different variance terms  $V(\varepsilon_i^E | \mathbf{x}_i^t)$  and  $V(\varepsilon_i^{WVS} | \mathbf{x}_i^t)$ . One such omitted factor is subjects' history of traumatic experiences which has been shown to be correlated with trust (Alesina and La Ferrara, 2002), but which is not observed by GLSS nor the present study. We investigated the amount of correlation between  $\varepsilon_i^E$  and  $\varepsilon_i^{WVS}$  in our data by jointly estimating equations (2) and (3), conditioning on the set of background characteristics corresponding to specification 4 in table 2.<sup>13</sup> The estimated value of  $Cor(\varepsilon_i^E, \varepsilon_i^{WVS} | \mathbf{x}_i^t)$  was 14.8%, significant at the 1% level. To get an idea of the impact an omitted variable can have on the estimated correlation between unobserved components, we removed lifetime experience with trust (TRUSTEXP) from the list of regressors. As conjectured, the estimated correlation increased from 14.8% to 17.28%.<sup>14</sup>

Second, differences in experimental design will directly influence the predictive power of survey trust measures via  $\xi$ , the common component between both trust measures, and via  $V(\varepsilon_i^E | \mathbf{x}_i^t)$ , the variance of the experimental design. The common element shared by two trust measures  $\xi$  is expected to be higher when both measures are tailored to capture the same type of trust. The WVS question explicitly elicits trust towards strangers. In our experiment, senders and receivers were strangers as their identities were never disclosed to the other party. On the other hand, subjects in GLSS were not strangers, as their identity was deliberately revealed amongst subject pairs in order to investigate the role of social connections.<sup>15</sup> We suspect that the fact that subjects new each other served as a behavioral discipline device, with subjects acting more kindly than they would otherwise if paired against strangers. This may well account for the high concentration of offers around the maximal amount sent reported by GLSS. Because our experimental design makes us more likely to capture WVS type trust than the experimental design of GLSS, we expect  $\xi$  to be relatively higher in our experiment. Finally, if whether subjects knew each other or not influences the distri-

---

<sup>13</sup>Tables of results are available upon request.

<sup>14</sup>Additionally removing age as a regressor further increased the correlation to 18.5%.

<sup>15</sup>In GLSS, pairs either knew each other before the experiment or got to know each other by jointly filling in a social connection survey. This survey includes among other questions to report the number of all personal acquaintances whom they have in common.



bution of responses, this design feature will also affect the variance term  $V(\varepsilon_i^E | \mathbf{x}_i^t)$ . Data of GLSS show that the unconditional variance in amounts sent is low, with 71% of their senders sending the maximal amount. In our experiment, the variance in the amount sent is much more dispersed which, from (5), implies that we should indeed expect higher values of  $\hat{\xi}$ .<sup>16</sup>

The framework above demonstrates that the predictive power of survey trust measures is intimately linked to the information experimenters possess on their subjects, and the choice of the experimental design. Because most of the factors are to some extent under the experimenters' control, prediction is not a suitable basis to compare experimental and survey trust. As the relevant policy exercise concerns measuring the effect of changes in the background characteristics on trust and trustworthiness, it seems more interesting to compare both measures on the basis of whether or not they carry the same information on the determinants of trust. This is something on which prediction has little to say.

### 3.1.2 Informational content

Measuring trust and trustworthiness using experiments is difficult when the target group is a nation's population. A more accessible alternative consists of surveys, which are easily found for a handful of countries. The methodological question is whether or not the effects of background characteristics on trust can be measured equally well using either experimental or survey data. If the effects measured were the same, the relative accessibility of survey measures would weaken the need for conducting experiments with representative samples.

By observing subjects' decisions in the trust game and their answers to the WVS trust question, we are in the unique position to compare how different the inferences on the determinants of trust can be when researchers use a popular survey question rather than experimental methods to measure trust. Specification 4 in table 2 reports results

---

<sup>16</sup> $V(\varepsilon_i^E | \mathbf{x}_i^t)$  is also likely to differ across studies according to differences in the dimension of the choice space of players, the fictitious currency used in the experiment, the monetary endowments, the multiplier, whether the experiment is conducted in the lab, in the classroom, at home on paper or on a computer, and whether the strategy method is used or not.

from a probit regression of the answers of our senders to the WVS trust question on their background characteristics. The differences with the experimental estimates are quite remarkable. We find that catholics and protestants trust others less than individuals without religious beliefs. This is in sharp contrast to the results from the experimental data where religious effects were totally absent. The second major difference concerns the education pattern. We do not find any effect of education on survey trust while we have found that a significant inverted-U relationship related experimental trust to education. Another notable difference is the effect of reported lifetime experience on trust. The effect is positive and significant when using the survey trust measure while it has an insignificant impact on experimental trust.

Despite these differences, there are some notable similarities. The inverted-U shape effect of age on trust remains when using survey trust. The number of children, subjects income, gender, and work status also remain insignificant using the survey measure.

To test whether the differences across measures significantly outweigh the similarities, we compare all effects simultaneously. In appendix A.1, we propose a simple minimum distance test which compares the difference between all parameter estimates of the probit model and the ordered probit model. We computed the test statistic based on specifications 2 and 4 of table 2. The value of the test is 21.78, significant at the 10 percent level. This indicates that the differences driven by the changes in the religious and education effects are strong enough to reject the null hypothesis that the informational content of both trust measures is identical.

### **3.2 Explaining differences between both measures**

The preceding section has shown that effects of background characteristics on trust are not robust to the type of measure used. We already mentioned that these differences can be attributed to the fact that the WVS question is vague and hypothetical or that it is misreported at the individual level. Recent developments in econometrics give us ways to test and control for misreporting of a discrete endogenous variable. If the difference between both experimental and survey trust measures are entirely due to misreporting, there is scope for using the WVS question and appropriately correct for misreporting. Furthermore, because our senders answered the survey question after

having played the trust game, there is the possibility that senders may have given answers to the WVS question which looked coherent with their play in the game. This would add to the amount of misreporting in our answers to the WVS question. This section tests for general misreporting and misreporting due to the sequence of play.

We define  $\alpha_{10}(y_i^E)$  as the probability that a subject answers 1 ("YES") to the WVS trust question when his truthful answer would be 0 ("NO"). This event occurs when, for example, generous senders are more inclined to state that they trust others rather than truthfully answering that they don't. Similarly, we define  $\alpha_{01}(y_i^E)$  as the probability that subject answers 0 ("NO") to the WVS trust question when his truthful answer would be 1 ("YES"). This probability captures events such as senders who are more likely to state that they do not trust others when they would have answered the opposite, had they been truthful. We allow for the fact that the amount of misreporting may depend on the experimental decision  $y_i^E$  by assuming the probabilities have the following logit form

$$\begin{aligned}\alpha_{10}(y_i^E) &= \frac{\exp(\theta_0^{10} + \theta_1^{10}y_i^E)}{\exp(\theta_0^{10} + \theta_1^{10}y_i^E) + 1} \\ \alpha_{01}(y_i^E) &= \frac{\exp(\theta_0^{01} + \theta_1^{01}y_i^E)}{\exp(\theta_0^{01} + \theta_1^{01}y_i^E) + 1}\end{aligned}$$

where  $\{\theta_0^{10}, \theta_0^{01}, \theta_1^{10}, \theta_1^{01}\}$  are unknown parameters to be estimated. Some special cases are of interest. If  $\theta_1^{10} = \theta_1^{01} = 0$  and the constant terms  $\theta_0^{10}$  and  $\theta_0^{01}$  are large, misreporting is random in the population of senders and is not affected by the preceding experiment. If in addition  $\theta_0^{10}$  and  $\theta_0^{01}$  are small, the misreporting probabilities are small indicating that senders truthfully answer the WVS question. Incorporation of these probabilities in a likelihood equation is a straightforward application of the results of Hausman, Abrevaya, and Scott-Morton (1998) and can be found in appendix A.2 of our paper.

Specification 5 of table 2 presents regression results for the probit model with misreporting. Both  $\theta_1^{10}$  and  $\theta_1^{01}$  associated with senders' experimental decisions  $y_i^E$  are not significantly different from zero, indicating that the experimental decision did not lead senders to systematically misreport their true answer to the WVS trust question. To test for overall misreporting, we computed a log-likelihood ratio test comparing

the probit model with and without misreporting (specifications 4 and 5).<sup>17</sup> The test value of 5.38 with 10% critical value of 7.02 does not reject the null hypothesis of no misreporting. Both sets of results suggest that senders in our experiment truthfully answered the WVS trust question, which means that allowing for the possibility that subjects misreport their answers is not sufficient to reconcile both measures. This suggests that the WVS trust question captures a notion of trust which differs from the notion of trust captured in our economic experiment.

## 4 Results on Trustworthiness

We follow GLSS and measure trustworthiness as the return ratio, defined as the amount returned divided by the amount available to return. In our experiment, the available amount to return equals the amount received multiplied by two, plus the experimental endowment of 500 CentERpoints. Responders were asked to play the strategy method by which they decide how much they will give back for each of the 11 possible amounts they can receive from the sender. This implies that we observe a sequence  $\{y_a^R \in [0, 1] \mid a \in \{0, 50, \dots, 500\}\}$  for each responder, where  $y_a^R$  denotes the return ratio when receiving  $a$  CentERpoints from the sender. The main advantage of the return ratio is that it is automatically scaled, which controls for the fact that receivers can send more simply because the total available amount increases with  $a$ . Figure 2 shows a boxplot of the return ratio for all 11 possible amounts. The thick line inside each box represents the median, the top and bottom of the boxes represent the 75th and 25th percentiles while the top and lower whiskers represent respectively the maximal and minimal values of the distribution.<sup>18</sup> The two important features of this figure are that the ratio 1) monotonically increases and is concave in the amounts received, 2) a significant fraction of the responders return nothing (especially in low categories) while practically no responder returns the entire possible amount.

Did it pay to trust? Figure 3 presents density estimates of senders returns of

---

<sup>17</sup>We use the test proposed by Andrews (2001) to deal with the fact that the probabilities of misreporting rest on the boundary of the parameter space under the null hypothesis. Computation of the test requires that the critical values of the log-likelihood ratio test be simulated. We report critical values based on 1000 simulations. See Andrews for further details.

<sup>18</sup>The minimal and maximal values are corrected to the presence of extreme outliers.

investments in trust.<sup>19</sup> Each line represents the estimated distribution of returns for a given number of CentERpoints sent. If responders return to senders exactly what they sent, the return on investment is 0. If responders do not return anything to the senders, the return is -1. Apart from the distribution of returns when 50 CentERpoints are sent, all distributions have roughly the same shape. The common finding in laboratory trust experiments is that trust barely pays, as responders return to senders what they have sent (Camerer, 2003). Our results reconfirm these findings. We find that the median return on investment is slightly above 0 for every amount sent. Furthermore, the probability of getting nothing back from a receiver (a return ratio of -1) is not zero.

The individual level analysis of the return ratio is based on the following Tobit model (Amemiya, 1984)

$$y_{ai}^{*R} = \mathbf{x}_i^{\mathbf{r}'}\eta + \gamma_1 a + \gamma_2 a^2 + \varepsilon_i^{\mathbf{r}} \quad i = 1, 2, \dots, N \quad (6)$$

$$y_{ai}^R = y_{ai}^{*R} \quad \text{if } y_{ai}^{*R} > 0 \quad (7)$$

$$= 0 \quad \text{if } y_{ai}^{*R} \leq 0 \quad (8)$$

Equation (6) describes the unconstrained trustworthiness propensity of responders. This propensity is modelled as depending on a vector of background characteristics  $\mathbf{x}_i^{\mathbf{r}}$ , an unobservable component  $\varepsilon_i^{\mathbf{r}}$ , and a vector of unknown population parameters  $\eta$ . The quadratic form in  $a$  is added to capture the increasing monotonicity in amounts returned observed in the data.<sup>20</sup> Equations (7) and (8) describe the censoring rule which allows responders with extremely low trustworthiness propensities to return nothing with positive probability.

The estimation results are presented in table 3. The first specification includes standard background characteristics of the responder, reported trust experience, their beliefs about what they expect to receive from the sender (RTHINK),<sup>21</sup> and responders' answers to the WVS trust question. The second specification removes the WVS trust

---

<sup>19</sup>These returns are computed as the (amount received - amount sent) / amount sent.

<sup>20</sup>We have estimated a less restrictive specification with dummy variables for each  $a$  category. Results were numerically identical to those presented above.

<sup>21</sup>RTHINK is coded from 0 to 10, where each unit is worth 50 CentERpoints.

answers while the third specification adds answers to the trust question and responders' beliefs. We compare the first three specifications using log-likelihood ratio tests. The extended specification which includes interaction terms is clearly preferred to the first two specifications.<sup>22</sup> Accordingly, our analysis below will focus on the results of the extended specification.

As could be seen from the raw data in figure 2, amounts returned monotonically increase and are concave in  $a$ , the amounts received. This is also reflected in the Tobit estimates, where the first order term  $\gamma_1$  is positive and the second order term  $\gamma_2$  is negative, both significant at the 1% level. The life cycle evolution of reciprocity is captured by the parameters of RETIRED, AGE, AGESQ, and the three interaction terms. The change in the trustworthiness propensity which follows from a change in age is given by (standard errors in parenthesis)

$$\frac{\partial R_{ia}^*}{\partial \text{AGE}_i} = 2 \times \underset{(0.000017)}{0.000032} \text{AGE}_i - \underset{(0.0005)}{0.0014} \text{WVS}_i - \underset{(0.0000)}{0.0003} \text{RTHINK}_i \quad (9)$$

We first solve the for turning points, which we define as the age around which the sign of the derivative changes, and compare them with those of trust found in the previous section. Because of the interaction terms, computation of the turning points requires that we fix the values of WVS and RTHINK. We can get an overall picture by evaluating equation (9) at the sample means of WVS and RTHINK. We find that trustworthiness reaches its lowest level when individuals reach the age of 30 years, and increases beyond that. These results differ remarkably from the life cycle evolution of trust discussed in the previous section. There, we found that trust increases until the age of 30 and decreases beyond that. The last section of the paper discusses a possible explanation of this result. We next evaluated the age turning points for those who report trusting others (WVS=1) and those who do not (WVS=0). The age profile of individuals who state they do not trust others reaches a low at 21 years of age, while it reaches a low at 43 years of age for those who declare trusting others.

In the previous section, we found that the relationship between trust and education was inverted U shape, with subjects without a secondary degree and those with

---

<sup>22</sup>The log-likelihood ration test value is 19.9 when comparing specifications 1 against 3, and 59.14 when comparing specifications 2 and 3. Both are significant at the 1% level.

university degrees displaying relatively less trustful behavior. The relation between education and trustworthiness is very different. Less educated subjects (the omitted category) return significantly more than educated subjects, all degrees confounded. Moreover, the parameter estimates suggests a U shape relationship, with individuals with university degrees being more trustworthy than subjects with technical education degrees.

The effect of gender also distinguishes trust from trustworthiness. While gender was found to have no impact on trust, we find here that men return on average significantly less than women.

One of the interesting findings of GLSS was that answers to the WVS trust question did not correlate with experimental trust but correlated rather well with the return ratio. In our experiment, we also find that subjects who trust more others are also more trustworthy. However, some of the issues raised in section 3.1.1 concerning the predictive power of stated trust questions may also apply here, although it is less clear from the results in this section whether the underlying process determining trustworthiness and trust are as similar.

We have shown in the previous section that subjects' beliefs were important determinants of trust. Here, beliefs of responders also play an important role in determining trustworthiness. Responders who believed they would receive more had higher average return ratios. To gain some insights on the importance of beliefs on trustworthiness, we estimated our extended specification omitting beliefs. Specification 4 in table 3 shows the result of this regression. The only notable change is that trustworthiness of those who report trusting others continues to decline with age while it no longer declines for those who report not trusting others. A log-likelihood ratio test (value of 85.88, significant at the 1% level) confirms that omitting beliefs substantially lowers the predictive fit of the model.

We end by noting that some individual characteristics have no effect on trustworthiness. This is the case of subjects' income, whether they work or not, their retirement status, religion, and their lifetime trust experience. Interestingly, none of these characteristics were found to explain experimental trust.

It is well known that the Tobit model is sensitive to the distributional assumption placed on the unobserved component (Newey, 1987). An alternative estimator which relaxes most distributional assumptions of the Tobit model is the Symmetrically Trimmed Least Squares estimator (STLS) of Powell (1986). Contrary to Tobit, the semiparametric STLS estimator does not require normality and is robust to (bounded) heteroscedasticity of unknown form in  $\varepsilon_i^r$ . All our specifications were estimated using the STLS estimator (results are presented in table 4). Hausman (1978) specification tests never reject the null hypothesis of normality and homoscedasticity of the error terms on which the Tobit model rests.

## 5 Results on Participation

The major impediment experimenters must overcome to test for participation bias is that they generally do not observe non-participants. In our experiment, we observe both the decision to participate and the characteristics of non-participants. This allows us to address the following issues 1) whether observed or unobserved factors are more important determinants of participation and 2) if participation is based on unobservable characteristics, are these related to experimental outcomes? The first question addresses the current belief in experimental economics which suggests that unobserved factors such as preferences for risk and money are more likely to explain participation in the experiment than the observed characteristics of subjects (Camerer and Hogarth, 1999). The second question directly touches the issue of participation bias, by which actions of participants are not representative of the population at large.

The most natural framework to approach both questions is that developed by Heckman (1978). We model participation as being driven both by observed and unobserved factors, the latter having the potential to affect the outcomes of interest and cause participation bias. Let  $d_i \in \{0, 1\}$  be an indicator of participation in the experiment and let an individual's unobserved latent propensity to participate be

$$d_i^* = \mathbf{x}_i^j \delta + \theta \text{RATIO}_i + \varepsilon_i^d \quad \text{for } j = \mathbf{r}, \mathbf{t}$$

where  $\mathbf{x}_i^j$  is the conditioning vector entering the trust and reciprocity models,  $\varepsilon_i^d$  is an unobservable determinant of participation assumed to be drawn from a  $N(0, 1)$



distribution, and  $(\delta, \theta)$  are unknown parameters. A general feature of these models is the requirement of a valid exclusion restriction for nonparametric identification of the participation bias. In practical terms, we need a variable which affects participation but does not directly affect either experimental and survey measures used in this paper. Our exclusion restriction is the variable `RATIO`, which is computed as proportion of questionnaires completed by panel members in the three months which preceded our experiment. This variable directly measures the participation propensity of subjects when participation is uncorrelated with financial outcomes.<sup>23</sup> The dependence between the experimental outcomes and the participation decisions is captured by the amount of correlation between  $\varepsilon_i^d$  and the unobservable components determining trust (see equation (1)) and trustworthiness (see equation (6)). We replicated the estimations of sections 2 and 4 by separately estimating an ordered probit, a binary probit model, and a Tobit model, along with the participation decision. We allow for potential participation bias by letting  $\varepsilon_i^d$  be correlated with the unobserved component of each experimental decision.

Most of the parameters entering the systematic part of the participation propensity were insignificant, confirming the conjecture that participation is mostly explained by unobserved characteristics of subjects.<sup>24</sup> One notable exception was income which has a positive and significant effect on participation, which rules out participation based on low opportunity costs. The presence of participation bias can be determined by testing the statistical significance of the correlation coefficients between  $\varepsilon_i^d$  and the unobserved components of the trust and trustworthiness measures used in this paper. We find that none of the three correlations are significant at the 10% level, a clear indication that the unobserved characteristics determining participation in the experiment do not correlate with the experimental decisions.

---

<sup>23</sup>CP members are not paid to participate in the panel.

<sup>24</sup>Results for trust and trustworthiness are virtually identical to those reported in the text. All but one parameter entering the trust propensity was significant (see text above). Results are available upon request.

## 6 Discussion and conclusions

The literature has identified trust and trustworthiness as important factors of economic performance and growth. Understanding the determinants of these at the societal level is important yet, not well documented. The majority of the existing empirical evidence relies on one of two complementary methodologies. Survey methods on one hand collect responses of heterogeneous samples, at the expense of having to rely on hypothetical and self-reported measures. On the other hand laboratory experiments offer the possibility to collect data on the actual behavior of subjects at the expense of collecting this data for a very special subset of the population of interest.

This paper presented results from a computerized experiment whose participants were randomly drawn from the Dutch population. This approach allowed us to combine the strengths of experiments and survey data collection methods.

One of the key findings of this paper is that background characteristics of subjects, mainly their age and education levels, do play an important role in determining trust and trustworthiness, although they affect trust and trustworthiness in very different ways. Our results reconfirmed the existing inverted U shape relation between trust and age, with trust increasing until the age of 30 and decreasing beyond that. On the other hand, we find that the relation between trustworthiness and age is U shaped, with trustworthiness decreasing until the age of 30, and rising again beyond that point. This raises the question of why do the young and elderly trust less but are more trustworthy than middle aged individuals? One explanation is that individuals who trust the most take for granted that the average individual in society will do the same. Hence, when someone places trust in them, they are less likely to be surprised by this action and will not place a premium as high on rewarding trust as would individuals who trust others less.

The education patterns are also very different. We find an inverted-U shape relation between education and trust, and an U shape relationship between education and trustworthiness. Such opposite influences of background characteristics are puzzling insofar as it is typically assumed that trust and trustworthiness go hand in hand, which would suggest that both are determined in similar ways. Reconciling the age

and education patterns of trust and trustworthiness is an interesting topic for future research.

An additional contribution of the paper is that we provided a new way to compare experimental and survey trust measures. The literature has up till now assessed the validity of survey trust questions by testing whether or not they predicted well experimental trust. One of the main messages of the paper is that this method of validation has been given too much attention, primarily because the predictive power of survey measures is intimately linked to the sample used, the amount of background information available on the subjects, and the experimental design. Our analysis has shown that by carefully selecting samples and designs, experimenters increase their odds of finding either a low or high predictive power of the survey trust measure. Thus, despite that contrary to the existing literature our survey trust measure predicts well trust in our experiment, we do not take this as evidence validating the use of survey trust questions. It is important to note that the problems with prediction are only relevant if predicting experimental trust with survey trust is what experimenters are trying to achieve. In general, prediction is useful if applied to an object which has a clear interest in being predicted well. Trust measured in an experiment is an abstract quantity whose main purpose is to extract from it useful characteristics of the population under study. Viewed in this way, it is not clear whether the emphasis on predicting its value by other trust measures is in general warranted.

As our paper focuses on investigating the determinants of trust and trustworthiness, it follows quite naturally that a more convincing comparison of both measures could be achieved by comparing whether experimental and survey measures have the same informational content. On these grounds, the differences between both trust measures are stark. We found that education has an inverted-U shape relation with experimental trust while it does not correlate at all with answers to the WVS trust question. In contrast, religion correlates well with answers to the WVS trust question but not at all with experimental trust.

We have also examined possible explanations for the differences between both measures. The two prominent explanations are that subjects either misreport their answers to the WVS question, or that the question simply captures a different notion of trust

than the one which is captured in the experiment. We do not find any evidence suggesting that these differences are due to subjects misreporting their answers to the survey trust question. As surveys remain more accessible than nation-wide laboratory experiments, it seems worthwhile for researchers interested in making cross-country comparisons to design new survey questions which will narrow the gap with experimental measures.

Finally, this paper made one of the first tests of participation bias in an economic experiment. We have not found any evidence suggesting the presence of participation bias in our experiment. In our view, this is a reassuring finding for experimental economics.

# A Econometric appendix

## A.1 Minimum distance test

Define  $\sigma^{WVS}$  and  $\sigma^E$  respectively as the standard deviation of  $\varepsilon^{WVS}$  and  $\varepsilon^E$ . Under the assumptions of the parametric models, the probit estimator provides consistent estimates of  $\beta^E/\sigma^E$  while the ordered probit model provides consistent estimates of  $\beta^{WVS}/\sigma^{WVS}$ . The comparison of both ratios is complicated by the fact that the experimental variance is partly under experimental control (see discussion in section 3.2). It is possible to get ride of  $\sigma^E$  by normalizing say the  $k$ th component to 1, which is equivalent to dividing the parameter vector by the  $k$ th component  $\beta_k^E/\sigma^E$ . Under this normalization,<sup>25</sup> the following transformed vector  $\left\{ \frac{\beta_1^E}{\beta_k^E}, \dots, \frac{\beta_{k-1}^E}{\beta_k^E}, 1, \frac{\beta_{k+1}^E}{\beta_k^E}, \dots, \frac{\beta_K^E}{\beta_k^E} \right\}$  is independent of  $\sigma^E$ . One can perform a similar division for parameter estimates from the probit model on the WVS trust question and obtain a second vector of parameters, this time independent of  $\sigma^{WVS}$ . In what follows, we use  $\xi^E$  and  $\xi^{WVS}$  as shorthand notations for the vectors of the ordered and binary probit models excluding the constant and the normalized  $k$ th term.<sup>26</sup> Under the null hypothesis that the effects of background characteristics are the same, both sets of estimates would equal each other.

Our test statistic has the following familiar quadratic form

$$N \left( \widehat{\xi}^E - \widehat{\xi}^{WVS} \right)' \mathbf{W}^{-1} \left( \widehat{\xi}^E - \widehat{\xi}^{WVS} \right) \quad (10)$$

where  $W$  represents the covariance matrix of the difference between both parameter vectors. Below we show that the test statistic above follows a chi square distribution with  $K - 2$  degrees of freedom.

We briefly sketch here a minimum distance test for observable characteristics. We will use  $\mathbf{E}_N$  to denote sample expectations and  $\mathbf{E}$  for corresponding population expectations. The estimated parameters solve

$$\begin{aligned} \widehat{\xi}^E &= \arg \max_{\xi} \mathbf{E}_N [L^E(\xi)] \\ \widehat{\xi}^{WVS} &= \arg \max_{\xi} \mathbf{E}_N [L^{WVS}(\xi)] \end{aligned}$$

---

<sup>25</sup>This normalization requires that  $\beta_k \neq 0$

<sup>26</sup>The constant term parameter is generally not separately identified from the threshold parameters in both the binary and ordered probit models. Given their values are functions of ad hoc threshold assumptions, they are not used in computation of the test.

where  $L^{WVS}(\xi)$  denotes the binary probit likelihood function and  $L^E(\xi)$  denotes the likelihood function of the ordered probit where the threshold parameters have been concentrated out. We will use the notation  $L_\xi(\xi)$  and  $L_{\xi\xi'}(\xi)$  to respectively denote the scores and the matrix of second derivatives of each function. It follows that

$$\begin{aligned} \sqrt{N} \begin{pmatrix} \widehat{\xi}^E - \xi_0 \\ \widehat{\xi}^{WVS} - \xi_0 \end{pmatrix} &= - \left\{ \begin{pmatrix} \left\{ \mathbf{E}_N [L_{\xi\xi'}^E(\bar{\xi})] \right\}^{-1} & \mathbf{0} \\ \mathbf{0} & \left\{ \mathbf{E}_N [L_{\xi\xi'}^{WVS}(\bar{\xi})] \right\}^{-1} \end{pmatrix} \right. \\ &\quad \left. \sqrt{N} \begin{pmatrix} \mathbf{E}_N [L_\xi^E(\xi_0)] \\ \mathbf{E}_N [L_\xi^{WVS}(\xi_0)] \end{pmatrix} \right\} + o_p(1) \end{aligned} \quad (11)$$

where every element of  $\bar{\xi}$  lies in between corresponding elements of  $\widehat{\xi}^E$  and  $\xi_0$ . Under appropriate regularity conditions (Newey and McFadden, 1994), Both  $\mathbf{E} [L_\xi^E(\xi_0)]$  and  $\mathbf{E} [L_\xi^{WVS}(\xi_0)]$  are zero vectors and

$$\begin{aligned} \mathbf{E}_N [L_{\xi\xi'}^E(\bar{\xi})] &\xrightarrow{U} \mathbf{E} [L_{\xi\xi'}^E(\xi_0)] \equiv \mathbf{H}_1 \\ \mathbf{E}_N [L_{\xi\xi'}^{WVS}(\bar{\xi})] &\xrightarrow{U} \mathbf{E} [L_{\xi\xi'}^{WVS}(\xi_0)] \equiv \mathbf{H}_2 \end{aligned}$$

where  $\xrightarrow{U}$  denotes uniform convergence. It follows that (11) converges in distribution to

$$\begin{aligned} &\mathcal{N} \left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}; \begin{pmatrix} \mathbf{H}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{J}^{E,E} & \mathbf{J}^{E,WVS} \\ \mathbf{J}^{WVS,E} & \mathbf{J}^{WVS,WVS} \end{pmatrix} \begin{pmatrix} \mathbf{H}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2^{-1} \end{pmatrix} \right) \\ &= \mathcal{N} \left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}; \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} \right) \end{aligned}$$

where  $J^{i,j} = \mathbf{E} [L_\xi^i(\xi_0) L_\xi^j(\xi_0)]$  for  $i, j = E, WVS$ . Finally, we get

$$\sqrt{N} (\widehat{\xi}^E - \widehat{\xi}^{WVS}) \xrightarrow{d} N(\mathbf{0}; \mathbf{W})$$

where  $\mathbf{W} \equiv \mathbf{V}_{11} + \mathbf{V}_{22} - \mathbf{V}_{12} - \mathbf{V}_{21}$ . Hence, equation (10) follows a chi square distribution with  $K - 2$  degrees of freedom. We compute the test statistic (10) by replacing  $\mathbf{W}$  with a sample average evaluated at either  $\widehat{\xi}^E$  or  $\widehat{\xi}^{WVS}$ , both equivalent under the null hypothesis.

## A.2 Probit model with misclassification

Following Hausman, Abrevaya, and Scott-Morton (1998), the probability that a subject states that he trusts others is given by

$$\Pr(y_i^{WVS} = 1 | \mathbf{x}_i^t, y_i^E) = \alpha_{10}(y_i^E) + (1 - \alpha_{10}(y_i^E) - \alpha_{01}(y_i^E)) \cdot \Pr(T_i^* > 0 | \mathbf{x}_i^t) \quad (12)$$

where  $(\alpha_{01}(y_i^E), \alpha_{10}(y_i^E))$  and  $T_i^*$  are defined in section 3. It follows that in the absence of misreporting errors ( $\alpha_{01}(y_i^E) = \alpha_{10}(y_i^E) = 0$ ), equation (12) collapses to standard binary probit model (Maddala, 1983). The likelihood function of the binary choice model with misreporting is constructed using the choice probabilities (12).

## B Instructions

The first 3 screens of the experiment are the same for both senders and responders. Italic notes in the translation are comments by the authors.

- *First screen:*

This experiment is a research project of researchers from \*\*\* and \*\*\*.

With this experiment you can make **real money** in terms of CentERpoints. You receive from the researchers **additional** CentERpoints (besides the usual telephone allowance).

- *Second screen:*

During this experiment you will be matched with another member of the panel. You will not know who this person is, **both of you will stay anonymous**. Both of you receive 500 CentERpoints. Then the experiments starts.

One of you has the possibility to send a share of this away. The amount of points sent will be doubled and given to the other person. The other person has then the opportunity to send a share of the own total amount back. The amount which is sent back will not be doubled.

How many points you finally earn depends therefore on your decision and the decision of the person you are matched with. You will be randomly assigned to your role.

- *Third screen:*

We now give you the chance to indicate whether you want to participate. If you decide not to participate, the experiment will end immediately. You will receive the usual telephone reimbursement. If you continue you will receive the 500 CentERpoints.

Do you want to continue?

- ☐ Yes  
☐ No

*Subjects who choose to participate were then randomly assigned to their roles. Senders and receivers had to read decision screens tailored to their roles.*

*Senders*

- *Fourth screen:*

You have been matched with another member of the panel. Like you, this person received 500 CentERpoints. You can send a share of your 500 CentERpoints. The panel member with whom you are matched with receives the amount you sent multiplied by 2. Then, this person has the opportunity to send a share of the own total amount back (without knowing who you are). The amount which this person sends back to you will not be doubled.

How many points do you want to give?

*(The sender could send one out of 11 possible amounts.)*

- ☐ **0** the other person receives additionally nothing and has therefore 500 and you remain with 500 points.

...

- ☐ **500** the other person receives additionally 1000 and has therefore 1500 in total and you remain with 0 points.

- *Fifth screen:*

*(was depending on the decision taken at the fourth screen, here as example "200")*

You decided to send **200** points.

The panel member you are matched with receives therefore 400 additional CentERpoints.



He or she has therefore in total **900** CentERpoints.

You remain with **300** CentERpoints.

How many points do you think the other panel member with whom you are matched with will send to you?

*(Participants had to type in a number. In this example in the range of  $[0,900]$ .)*

- *Sixth screen:*

This experiment is done with some panel members. Half of them interact in the same position as you. They can send a share of their 500 CentERpoints which is doubled and received by a person of the other position.

How many points do you think those panel members have sent?

*(The sender could indicate one out of 11 possible amounts from 0 to 500).*

### *Responders*

- *Fourth screen:*

You have been matched with another member of the panel. Like you, this person received 500 CentERpoints. This person is asked to send you a share from their own 500 CentERpoints. You will receive the amount of those points the other person has sent multiplied by 2.

For example, if the other person sends 100 CentERpoints, you will receive 200 CentERpoints. Together with the 500 points you begin with, you will have in total 700 CentERpoints.

From this amount you can return a share. The amount you send will not be doubled.

- *Fifth screen:*

As we do not know now how many CentERpoints the other panel member with whom you are matched with has sent we present all possible amounts this person could send to you. The amount you actually receive is written in the next column. Please indicate in the last column what amount you would return **for each possible amount sent**.

(The responder had to indicate for each of 11 possible amounts the sender could send what he would return. The table was designed as follows:)

0	0	500	-----
...	...	...	...
500	1000	1500	-----

How many points do you **expect** the panel member with whom you are matched with has sent to you?

After these screens the experiment was over. Nobody could go backwards and both senders and responders were asked the following post-experimental questions:

The last two questions are about trust in general. This question is about your own trust experience.

Choose the number which is closest to your answer.

(Participants had to type in a number between 1 and 7).

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

- 33

	<i>S</i>		<i>R</i>		Not played		
	Mean	StD	Mean	StD	Mean	StD	
GENDER	0.46	0.5	0.46	0.5	0.43	0.5	Dummy, 1 if Man
AGE	42.66	15.55	43.98	15.67	53.79	16.55	Age
GR_INCP	2580.38	6133.47	1916.28	3289.66	2356.90	3584.27	Gross Personal Monthly Income (Guilders)
GR_INCH	4175.59	7526.39	3342.07	4313.76	4302.52	6593.33	Gross Household Monthly Income (Guilders)
HSIZE	2.27	1.12	2.53	1.33	2.17	1.23	Household Size
NCHILD	0.55	0.92	0.75	1.16	0.4	1.06	Number of Children in household
PARTNER	0.72	0.45	0.78	0.42	0.76	0.43	Partner is living in the household
LOWEDUC	0.043	0.204	0.076	0.266	0.143	0.354	Dummy, 1 if Low education level
SECONDEG	0.388	0.488	0.390	0.489	0.381	0.492	Dummy, 1 if Secondary degree
TRAINDEG	0.449	0.449	0.448	0.499	0.357	0.485	Dummy, 1 if Training degree
UNIVDEG	0.119	0.325	0.085	0.279	0.119	0.329	Dummy, 1 if University degree
WORK	0.591	0.493	0.511	0.501	0.429	0.501	Dummy, 1 if Labor Work
RETIRED	0.130	0.337	0.148	0.356	0.262	0.445	Dummy, 1 if Retired
HHEAD	0.637	0.482	0.609	0.488	0.667	0.477	Dummy, 1 if Head of household
CATHOLIC	0.297	0.458	0.349	0.478	0.309	0.468	Dummy, 1 if Catholic religion
PROTEST	0.210	0.408	0.265	0.442	0.191	0.397	Dummy, 1 if Protestant religion
OTHERS	0.493	0.501	0.386	0.488	0.500	0.506	Dummy, 1 if Atheist or Other religion
WVS	0.533	0.500	0.488	0.498	—	—	Dummy, 1 if Yes to WVS trust question
TRUSTEXP	3.16	1.18	3.45	1.31	—	—	1-7, 1 if Always rewarded 7, Always disappointed
Median time*	3.29		5.93		1.28		Minutes taken to play the experiment
Nobs	276		223		42		

Table 1: Descriptive Statistics. \*Median time played is computed from the time subjects logged in the CentERdata network.

Specification	Experimental trust						Stated trust (WVS)			
	(1)		(2)		(3)		(4)		(5)	
	<i>Coef.</i>	<i>Std.</i>	<i>Coef.</i>	<i>Std.</i>	<i>Coef.</i>	<i>Std.</i>	<i>Coef.</i>	<i>Std.</i>	<i>Coef.</i>	<i>Std.</i>
CONSTANT	-2.018**	0.736	-0.314	0.681	-1.828***	0.687	-3.069***	0.838	-3.787	1.424
GENDER	0.182	0.133	0.229*	0.134	0.217	0.160	-0.186	0.181	-0.335	0.244
AGE	0.067**	0.031	0.052*	0.029	0.060***	0.031	0.067*	0.037	0.071	0.051
AGE_SQ/1000	-0.897**	0.367	-0.662**	0.335	-0.803**	0.349	-0.909**	0.435	-1.026*	0.589
RETIRED	-0.193	0.374	0.271	0.293	-0.251	0.287	0.528	0.411	0.707	0.535
SECONDEG	0.908***	0.396	0.626*	0.308	0.845***	0.318	0.518	0.422	0.504	0.578
TRAINDEG	0.809**	0.404	0.589*	0.312	0.761**	0.332	0.394	0.436	0.446	0.602
UNIVDEG	0.729	0.457	0.716*	0.361	0.678	0.395	0.522	0.499	0.337	0.678
WORK	-0.219	0.193	-0.089	0.185	-0.189	0.208	-0.290	0.243	-0.451	0.373
LN(GR.INCP)	0.024	0.026	0.007	0.028	0.019	0.039	0.034	0.038	0.044	0.053
NUMCHILD	-0.118*	0.064	-0.043	0.066	-0.129	0.099	0.105	0.102	0.121	0.146
CATHOLIC	0.101	0.170	0.002	0.165	0.149	0.171	-0.476**	0.211	-0.567*	0.290
PROTEST	-0.016	0.164	-0.109	0.149	0.042	0.208	-0.524**	0.227	-0.699**	0.355
TRUSTEXP	0.045	0.059	0.090	0.065	-0.032	0.082	0.665**	0.085	0.905***	0.205
STHINK	0.003***	0.0001			0.003***	0.000				
SMEANS	0.423***	0.061			0.433***	0.044				
WVS					0.362**	0.169				
$\theta_0^{10}$									-3.467	1.621
$\theta_1^{10}$									0.387	0.233
$\theta_0^{01}$									-1.311	1.881
$\theta_1^{01}$									-1.119	1.583
Log-Likelihood	-406.42		-520.06		-403.58		-150.8		-148.11	

Table 2: Sender Results. Specifications (1) to (3) refer to the ordered probit model, columns (4) and (5) to the binary probit model. The significance of parameters is based on robust standard errors and bootstrap empirical quantiles of the t-statistics (1000 repetitions), N=276.

	(1)		(2)		(3)		(4)	
	<i>Coef.</i>	<i>Std.</i>	<i>Coef.</i>	<i>Std.</i>	<i>Coef.</i>	<i>Std.</i>	<i>Coef.</i>	<i>Std.</i>
CONSTANT	-0.042	0.032	-0.071**	0.032	-0.076***	0.035	-0.056	0.035
$\gamma_1$	0.087***	0.004	0.0875***	0.004	0.087***	0.004	0.088***	0.004
$\gamma_2$	-0.005***	0.0003	-0.005***	0.0003	-0.0045***	0.0003	-0.005***	0.0003
GENDER	-0.018***	0.007	-0.019**	0.007	-0.017***	0.0068	-0.008	0.007
AGE	-0.002*	0.001	-0.001	0.001	-0.0014	0.0014	0.0003	0.001
AGESQ/1000	0.033***	0.017	0.019	0.016	0.032*	0.017	0.004	0.017
RETIRED	-0.004	0.015	-0.007	0.015	0.0052	0.014	0.017	0.015
SECONDEG	-0.053***	0.012	-0.060***	0.012	-0.054***	0.012	-0.052***	0.012
TRAINDEG	-0.065***	0.012	-0.067***	0.012	-0.067***	0.012	-0.062***	0.012
UNIVDEG	-0.053***	0.015	-0.057***	0.015	-0.058***	0.015	-0.058***	0.015
WORK	0.016*	0.009	0.011	0.009	0.019**	0.0089	0.013	0.009
LN(GR_INCP)	0.0006	0.0011	0.002	0.001	0.0001	0.0011	0.0002	0.001
CATHOLIC	-0.009	0.007	-0.009	0.007	-0.0062	0.0073	-0.002	0.007
PROTEST	0.006	0.008	0.004	0.008	0.0051	0.0076	-0.0005	0.008
WVS	0.045***	0.007			0.1035***	0.022	0.149***	0.021
TRUSTEXP	-0.003	0.003	0.0064**	0.0027	-0.017	0.009	-0.011	0.009
RTHINK	0.013***	0.001	0.013***	0.001	0.0255***	0.0044		
AGE $\times$ TRUSTEXP					0.0003	0.0002	0.0003	0.0002
AGE $\times$ WVS					-0.0014***	0.0005	-0.002***	0.0005
AGE $\times$ RTHINK					-0.0003***	0.0000		
$\sigma^2$	0.144***	0.002	0.021***	0.0006	0.0207***	0.0006	0.021***	0.0006
Log-Likelihood	936.52		916.75		946.47		903.53	

\*\*\*: 1% Level. \*\*: 5% Level. \*: 10% Level

Table 3: Responder Results - Tobit estimator. N = 2453.

	(1)		(2)		(3)		(4)	
	<i>Coef.</i>	<i>Std.</i>	<i>Coef.</i>	<i>Std.</i>	<i>Coef.</i>	<i>Std.</i>	<i>Coef.</i>	<i>Std.</i>
CONSTANT	-0.039	0.033	-0.069**	0.033	-0.042	0.036	-0.043	0.037
$\gamma_1$	0.079***	0.004	0.079***	0.005	0.079***	0.005	0.079***	0.005
$\gamma_2$	-0.004***	0.0004	-0.004***	0.0004	-0.004***	0.0004	-0.004***	0.0004
GENDER	-0.013**	0.007	-0.015**	0.007	-0.014**	0.007	-0.003	0.006
AGE	-0.001	0.001	0.0002	0.001	-0.0006	0.001	0.002	0.001
AGESQ/1000	0.017	0.017	0.004	0.016	0.009	0.018	-0.021	0.018
RETIRED	-0.002	0.017	-0.005	0.017	0.006	0.017	0.021	0.017
SECONDEG	-0.050***	0.016	-0.059***	0.016	-0.053***	0.015	-0.045**	0.016
TRAINDEG	-0.064***	0.016	-0.065***	0.016	-0.066***	0.015	-0.057**	0.016
UNIVDEG	-0.040***	0.018	-0.045***	0.018	-0.0469**	0.018	-0.043**	0.018
WORK	0.012	0.008	0.008	0.008	0.0128	0.008	0.008	0.008
LN(GR.INCP)	0.000	0.001	0.001	0.001	-0.0005	0.001	-0.0006	0.001
CATHOLIC	-0.014**	0.007	-0.012	0.007	-0.013	0.007	-0.008	0.007
PROTEST	0.010	0.007	0.009	0.007	0.012	0.008	0.004	0.008
WVS	0.042***	0.007			0.086***	0.020	0.129***	0.021
TRUSTEXP	-0.001	0.003	0.0082**	0.003	-0.025**	0.009	-0.017*	0.009
RTHINK	0.012***	0.001	0.012***	0.001	0.022***	0.0046		
AGE $\times$ TRUSTEXP					0.0006**	0.0002	0.0004**	0.0002
AGE $\times$ WVS					-0.0011**	0.0005	-0.002***	0.0005
AGE $\times$ RTHINK					-0.0002**	0.0000		
Hausman $\chi^2$ test	3.106		2.762		15.656		19.234	

\*\*\*: 1% Level. \*\*: 5% Level. \*: 10% Level

Table 4: Responder Results - Symmetrically Trimmed Least Squares estimator. N = 2453.

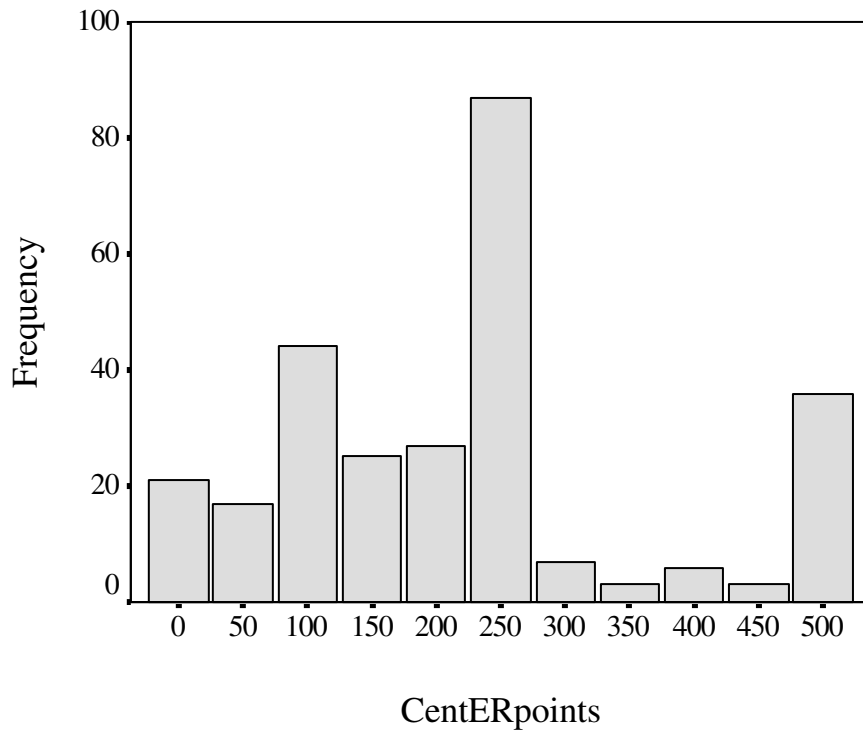


Figure 1: Distribution of experimental trust.

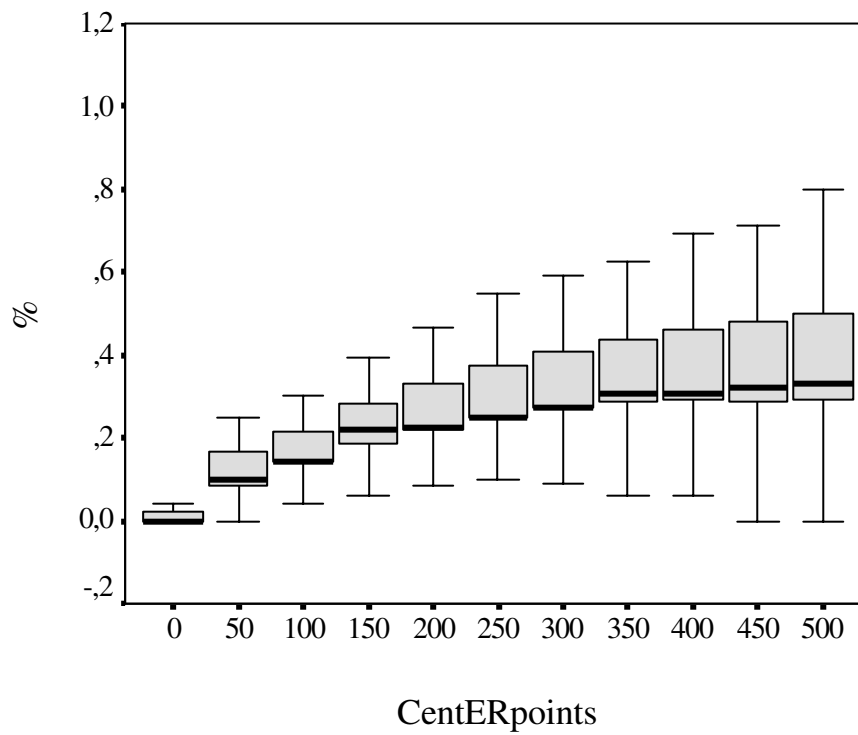


Figure 2: Return ratio of responders for each units received, strategy method.

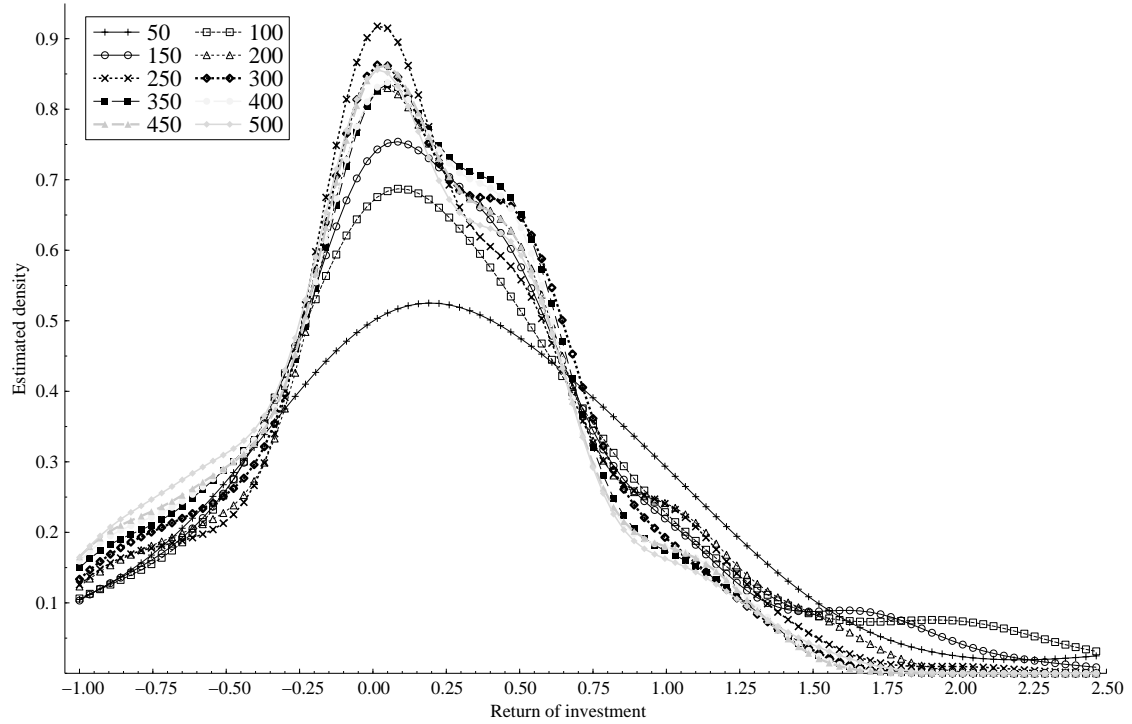


Figure 3: Estimated density of potential returns on investments in trust for each amount sent. Gaussian kernel density estimation. Rate of return computed as (amount returned - amount sent) / amount sent. The rate of return is infinity when the amount sent is zero and is not plotted here.

## References

- ALESINA, A., AND E. L. FERRARA (2002): “Who Trust Others,” *Journal of Public Economics*, 85(2), 207–234.
- AMEMIYA, T. (1984): “Tobit Models: A Survey,” *Journal of Econometrics*, 24, 3–61.
- ANDERHUB, V., D. ENGELMANN, AND W. GÜTH (1999): “An Experimental Study of the Repeated Trust Game with Incomplete Information,” Humboldt-Universität zu Berlin, Sonderforschungsbereich 373 Discussion Paper No.97.
- ANDREWS, D. W. K. (2001): “Testing When a Parameter is on the Boundary of the Maintained Hypothesis,” *Econometrica*, 69(3), 683–734.
- BERG, J., J. DICKHAUT, AND K. MCCABE (1995): “Trust, Reciprocity, and Social History,” *Games and Economic Behavior*, 10, 122–142.



- BOSCH-DOMÈNECH, A., J. G. MONTALVO, R. NAGEL, AND A. SATORRA (2002): “One, Two, (Three), Infinity, ... : Newspaper and Lab Beauty-Contest Experiments,” *American Economic Review*, 92(5), 1687–1701.
- BOUND, J., C. BROWN, AND N. MATHIOWETZ (2001): *Measurement Error in Survey Data* chap. 59, pp. 3705–3843. Elsevier.
- BOWLES, S., AND H. GINTIS (2002): “Social Capital and Community Governance,” *The Economic Journal*, 112(483), F419–F436.
- BRANDTS, J., AND G. CHARNESS (2000): “Hot Vs. Cold: Sequential Responses and Preference Stability in Experimental Games,” *Experimental Economics*, 2, 227–238.
- BURKS, S. V., J. P. CARPENTER, AND E. VERHOOGEN (2003): “Playing Both Roles in the Trust Game,” *Forthcoming, Journal of Economic Behavior and Organization*.
- CAMERER, C. (2003): *Behavioral Game Theory: Experiments on Strategic Interaction*. Princeton University Press, Princeton.
- CAMERER, C. F., AND R. M. HOGARTH (1999): “The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework,” *Journal of Risk and Uncertainty*, 19:1-3, 7 – 42.
- CROSON, R. (2000): “Thinking Like a Game Theorist: Factors Affecting the Frequency of Equilibrium Play,” *Journal of Economic Behavior and Organization*, 41, 299–314.
- DURLAUF, S. N. (2002): “On the Empirics of Social Capital,” *The Economic Journal*, 112(483), F459–F479.
- ECKEL, C., AND P. GROSSMAN (2000): “Volunteers and Pseudo-Volunteers: The Effect of Recruitment Method on Subjects’ Behavior in Experiments,” *Experimental Economics*, 3, 107–120.
- FEHR, E., U. FISCHBACHER, B. V. ROSENBLADT, J. SCHUPP, AND G. WAGNER (2002): “A Nation-Wide Laboratory,” *Schmollers Jahrbuch*, 122, 519–542.
- FRANCOIS, P., AND J. ZABOJNIK (2002): “Trust as Social Capital and the Process of Economic Development,” *Tilburg University*, p. mimeo.

- GLAESER, E. L., D. LAIBSON, AND B. SACERDOTE (2002): “An Economic Approach to Social Capital,” *The Economic Journal*, 112(483), F437–F458.
- GLAESER, E. L., D. LAIBSON, J. A. SCHEINKMAN, AND C. SOUTTER (2000): “Measuring Trust,” *Quarterly Journal of Economics*, 115(3), 811–846.
- GOTTSCHALK, P., AND T. M. SMEEDING (1997): “Cross-National Comparisons of Earnings and Income Inequality,” *Journal of Economic Literature*, 35, 633–687.
- GRUBER, J., AND D. WISE (2001): “An International Perspective on Policies for an Aging Society,” *NBER Working Paper*, 8130.
- HARRISON, G. W., M. I. LAU, AND M. B. WILIAMS (2002): “Estimating Individual Discount Rates in Denmark: A Field Experiment,” *American Economic Review*, 92(5), 1606–1617.
- HAUSMAN, J., J. ABREVAYA, AND F. SCOTT-MORTON (1998): “Misclassification of a Dependent Variable in a Discrete Response Setting,” *Journal of Econometrics*, 87, 239–269.
- HAUSMAN, J. A. (1978): “Specification Tests in Econometrics,” *Econometrica*, 46, 1251–1271.
- HECKMAN, J. J. (1978): “Sample Selection Bias as a Specification Error,” *Econometrica*, 47(1), 153–162.
- HEY, J. D. (2002): “Experimental Economics and the Theory of Decision Making under Risk and Uncertainty,” *Geneva Papers on Risk and Insurance Theory*, 27(1).
- KNACK, S., AND P. KEEFER (1997): “Does Social Capital Have an Economic Payoff? A Cross-Country Investigation,” *Quarterly Journal of Economics*, 112(4), 1252–1288.
- LAPORTA, R., F. L. DE SILANES, A. SHLEIFER, AND R. W. VISHNY (1997): “Trust in Large Organizations,” *American Economic Review*, 87(2), 333–338.

- LUCKING-REILEY, D. H. (1999): “Using Field Experiments to Test Equivalence Between Auction Formats: Magic on the Internet,” *American Economic Review*, 89(5), 1063–1080.
- MADDALA, G. (1983): *Limited Dependent and Qualitative Variables in Econometrics*, Econometric Society Monographs. Cambridge University Press, Cambridge, UK.
- NEWKEY, W. K. (1987): “Specification Tests for Distributional Assumptions in the Tobit Model,” *Journal of Econometrics*, 34, 125–145.
- NEWKEY, W. K., AND D. L. MCFADDEN (1994): *Large Sample Estimation and Hypothesis Testing* vol. 4 of *Handbook of Econometrics*, chap. 36, pp. 2111–2245. Elsevier Science B.V.
- ORTMANN, A., J. FITZGERALD, AND C. BOEING (2000): “Trust, Reciprocity, and Social History: A Re-Examination,” *Experimental Economics*, 3, 81–100.
- POWELL, J. L. (1986): “Symmetrically Trimmed Least Squares Estimation for Tobit Models,” *Econometrica*, 54(6), 1435–1460.
- PUTNAM, R. (1993): *Making Democracy Work*. Princeton University Press, Princeton.
- PUTNAM, R. (2000): *’Bowling Alone: The Collapse and Revival of the American Community*. Simon and Schuster, New-York.
- RUUD, P. A. (2000): *An Introduction to Classical Econometric Theory*. Oxford University Press, New-York.
- SELTEN, R., AND A. OCKENFELS (1998): “An Experimental Solidarity Game,” *Journal of Economic Behavior and Organization*, 34(4), 517–539.
- WILLINGER, M., C. KESER, C. LOHMANN, AND J.-C. USUNIER (2003): “A Comparison of Trust and Reciprocity Between France and Germany: Experimental Investigation Based in the Investment Game,” *Forthcoming, Journal of Economic Psychology*.
- ZAK, P. J., AND S. KNACK (2001): “Trust and Growth,” *Economic Journal*, 111(4), 295–321.