DISCUSSION PAPER SERIES

IZA DP No. 4087

Measuring Trust: Experiments and Surveys in Contrast and Combination

Michael Naef
Jürgen Schupp

March 2009

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Measuring Trust: Experiments and Surveys in Contrast and Combination

Michael Naef
Royal Holloway, University of London

Jürgen Schupp
SOEP, DIW Berlin and IZA

Discussion Paper No. 4087
March 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 4087
March 2009

ABSTRACT

Measuring Trust: Experiments and Surveys in Contrast and Combination*

Trust is a concept that has attracted - significant attention in economic theory and research within the last two decades: it has been applied in a number of contexts and has been investigated both as an explanatory and as a dependent variable. In this paper, we explore the questions of what exactly is measured by the diverse survey-derived scales and experiments claiming to measure trust, and how these different measures are related. Using nationally representative data, we test a commonly used experimental measure of trust for robustness to a number of interferences, finding it to be mostly unsusceptible to stake size, the extent of strategy space, the use of the strategy method, and the characteristics of the experimenters. Inspired by criticism of the widespread trust question used in many surveys, we created a new, improved survey trust scale consisting of three short statements. We show that the dimension of this scale is distinct from trust in institutions and trust in known others. Our new scale is a valid and reliable measure of trust in strangers. The scale is valid in the sense that it correlates with trusting behaviour in the experiment. Furthermore, we demonstrate that the test-retest reliability of six weeks is high. The experimental measure of trust is, on the other hand, not significantly correlated with trust in institutions nor with trust in known others. We therefore conclude that the experimental measure of trust refers not to trust in a general sense, but specifically to trust in strangers.

JEL Classification: C83, C91, D63, Z13

Keywords: trust, experiment, survey, representativity, SOEP

Corresponding author:

Jürgen Schupp
Socio-Economic Panel Study (SOEP)
DIW Berlin
D-10108 Berlin
Germany
E-mail: jschupp@diw.de

* We want to thank Ernst Fehr, Urs Fischbacher and Gert G. Wagner with who we designed the survey, for their contribution and helpful comments; Andreas Stocker and Bernhard von Rosenbladt for their support in finding practical solutions for the field work of the surveys; and Deborah Anne Bowen for her critical review of the manuscript.

1 Introduction

In surveys like the General Social Survey (GSS) or the World Values Survey (WVS), trust is measured with the statement "*Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?*" This measure of trust has been criticized and its behavioural relevance has been called into question.

The first systematic study of the relation between survey and behavioural measures of trust was reported by Glaeser et al. (2000). They investigated whether behaviour in a trust game is correlated with this standard survey measure of trust. They find that the above question is not correlated with trusting behaviour. This result has been replicated in several other studies (e.g., Ashraf et al. 2003, Ermisch et al. 2007, Gachter et al. 2004, Haile et al. 2008, Holm & Nystedt 2008, Johansson-Stenman et al. 2005b). However, other studies have found a significant correlation between the survey and the experimental measures (e.g., Vyrastekova & Garikipati 2005, Bellemare & Kroeger 2007, Sapienza et al. 2007). Based on the previous research one cannot conclude whether the GSS question is behaviourally relevant in the sense that it correlates with the behaviour in the trust game.

What are the reasons for these conflicting results? Are the experiment and the survey measures both valid and reliable measures of trust? Concerning survey measures, several studies have revealed that the GSS question is neither a valid nor a reliable measure of trust (Reeskens & Hooghe 2008). The question is rather imprecise, the possible answers are not mutually exclusive, and only one item is not considered to be a reliable measurement (e.g., Glaeser et al. 2000, Miller & Mitamura 2003, Yamagishi et al. 1999). Concerning the experimental measure, little is known about its sensitivity and validity in large and heterogeneous populations.[1] Against this background, it is no surprise that there is no clear relationship between survey and experimental measures of trust.

The aim of this article is to connect the survey measures and the experimental measures of trust. In particular, we would like to show that survey and experimental

[1] Most experiments are run with students as subjects

measures can be connected in a large representative survey. Since the experimental measure might capture a very specific dimension of trust, we created a new survey measure of trust that, on the one hand, takes recent criticisms of the GSS question into account, and on the other, measures the same dimension of trust as the experiment.

To avoid confusion we have to clarify first what we mean by the word "trust". We largely follow James Coleman's concept of trust (Coleman 1990). From his perspective, the following two points characterise the action of placing trust. On the one hand, trust implies that the truster freely transfers assets to another person, without controlling the actions of that other person or having the possibility to retaliate. On the other hand, there must be a potential gain in order to have an incentive to trust. The incentive is such that the truster is better off than not having trusted if the other person is trustworthy, and worse off if the other person does not merit the trust placed in him/her. Note that in this concept, trust is considered a form of behaviour rather than as a personal characteristic or personal trait.

Our new survey measure is more precise than the GSS question on what dimension of trust is measured. We focus on trust in strangers. We have used this new survey measure in the German Socio-Economic Panel (SOEP) study and several independent studies that are all representative for the German population. In order to distinguish the newly developed trust questions from the GSS question, we refer to them as "SOEP-trust". A factor analysis shows that SOEP-trust (trust in strangers) measures a different dimension of trust than questions on trust in institutions and trust in known others. Further, we show that SOEP-trust is a valid and a sensitive measure of trust. Concerning the latter, we find that SOEP-trust is correlated with social desirability and the position in the survey. Furthermore, trust is only moderately stable over three weeks. This has implications for the use of the survey question in international comparisons and over time.

The design of the simplified trust game is as follows. Two players are each endowed with 10 euros. The first mover decides how many of his or her 10 euros he or she would like to transfer to the second mover. Each transfer is doubled by the experimenters.

The second mover then gets to know the first mover's transfer and then decides him or herself about the back-transfer. As with the first mover, the second mover can transfer any amount between zero and ten euros. The second mover's transfer is doubled as well. Then the game is over and both participants are paid by a cheque. In order to distinguish the experiment from other trust measures we refer to it as "EXP-trust" in the following.

Our implementation of the experiment in a large survey has several advantages over an ordinary laboratory experiment. With our survey measure of trust, we are able to directly check whether only highly trusting people decide to participate in the trust game. Since this is a panel study, we can additionally check whether less trusting people are more likely to leave the panel in the future. Further, we can compare students and non-students in the same experimental setting (design and procedure).

Another advantage of the combination of survey and experiments is that we can use the information in the survey to validate EXP-trust. We thus analyse whether the decision to trust is influenced by risk preferences, selflessness, and expectations, as postulated by economic theory. We also assessed the sensitivity of the experimental design to a social desirability bias, the stake size, and the available strategy space.

We find that EXP-trust is surprisingly robust and also not subject to a social desirability bias, and not dependent on the exact stake size or on the size of the strategy space. Furthermore, we find that for subjects who are familiar with the interview situation (i.e., through previous participation in a panel study), selection into the experiment is not subject to their level of trust. In contrast, for subjects who have not previously been part of a panel study, more trusting people are more likely to participate in the experiment. And contrary to previous research, we find that students are more trusting than non-students, which has consequences for the generalisability of experimental results from students to the general population.

Finally we analyse what kind of trust the experiment actually measures. We find that EXP-trust measures people's trust in strangers, but not their trust in institutions or in known others. That is, EXP-trust is significantly correlated with the newly developed

SOEP-trust measure. Past trusting behaviour is a good predictor of the behaviour in the trust game as well.

Thus, on a representative level for Germany, we show that survey and experimental measures of trust are connected in the way that the trust game measures a specific dimension of trust, that is, trust in strangers.

In the following, we first analyse the survey measure of trust (SOEP-trust), and in the second part analyse the experimental trust measure (EXP-trust). In the third part, we combine these two measures and analyse their similarities and differences.

2 Using Surveys to Measure Trust

In this chapter, we propose a new measure of trust in strangers and analyse its sensitivity, reliability, and validity. In particular, we show that trust in strangers measures a specific dimension of trust that is distinct from other dimensions like trust in institutions or trust in known others. We implement this entire analysis in the framework of the German Socio-Economic Panel (SOEP). The SOEP is a household panel that contained 22,611 individuals in 12,061 households in the year 2003, comprising a representative sample for Germany (Wagner et al. 2007). Additionally, we conducted four accompanying studies (AS), one each year from 2002 to 2006. These accompanying studies all have a randomly drawn sample of 400 to 1,000 observations of the German population and are thus representative for Germany as well. Together these data sets provide a great tool to assess survey questions in a large heterogeneous population. In Table A.1 we list all the studies used and give an overview of which study we used to implement the different variations of the study design. In Appendix A.2, we discuss whether our results can be viewed as representative for trust.

2.1 SOEP-trust — A New Measure of Trust

The General Social Survey (GSS) measure of trust, together with the quite similar question in the World Values Survey (WVS)[2] is probably the most widely used question to measure trust in surveys.

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

- *Most people can be trusted*
- *Can´t be too careful*

This question measures people's expectations of others' trustworthiness. Based on our concept of trust, expectations about other people's trustworthiness is an important factor in deciding whether one decides to trust or distrust. The advantage of this question is that the same question is used over time and space, thus allowing a wide array of different analyses. However, it has been criticized that the respondents have the choice between trust and caution and not between trust and distrust or between cautious and incautious behaviour (for a review, see Yamagishi et al. 1999). Although trust and caution are difficult to disentangle, it is important to measure them separately, since trust and caution are not necessarily mutually exclusive. The interpretation of the GSS question can therefore differ widely among different societies (e.g., Gabriel et al. 2002). Miller & Mitamura (2003) showed, for example, that Japanese students are more trusting than Americans measured with the above question from the GSS. Measuring trust and caution separately, they find, however, that American students are more trusting than Japanese students but at the same time also more cautious. These differing results clearly demonstrate the problems for the interpretation of the above question.

Based on this evidence, we decided to create a new measure of trust using the German Socio-Economic Panel. We split the GSS question up into two parts. On the one hand, we asked people to what extent they agree with the following two statements:

[2]Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people? Most people can be trusted OR Need to be very careful.

- *In general, you can trust people.*
- *Nowadays, you can't rely on anybody.*

The possible answers on a four point rating scale were "disagree strongly", "disagree somewhat", "agree somewhat", or "agree strongly".

Another criticism of the GSS question is that answers may differ significantly depending on whether people understand "most people" in the question as meaning acquaintances or strangers (Reeskens & Hooghe 2008). We therefore let people rate their agreement with two further statements about trust and caution, in which we clearly state that trust towards strangers is meant and not towards family or friends:

- *How much do you trust strangers you meet for the first time.*
- *When dealing with strangers, it's better to be cautious before trusting them.*

The possible answers on a four-point rating scale were either "no trust at all", "little trust", "quite a bit of trust", and "a lot of trust" for the first question and "disagree strongly", "disagree somewhat", "agree somewhat", or "agree strongly" for the second question. These four statements constitute our new measure of trust in strangers. To distinguish it from GSS trust, we will call it SOEP-trust in the following.

The emphasis on trust in strangers takes into account that trust is a multidimensional concept. To test whether SOEP-trust measures trust in strangers specifically, we let people rate other statements on trust in different institutions like the police or the government and in acquaintances like friends and family. The list of items can be seen in Table 1. People could answer on the scale from "no trust at all", "little trust", "quite a bit of trust", to "a lot of trust". Because of the multidimensionality of trust, these items are expected to measure different aspects of trust. A principal component analysis over all the trust items can show us how many dimensions these items measure. We find that these items represent three distinct components.

Table 1 reports the factor loadings of the different items. The bold numbers indicate the component to which each item belongs. Each component has a straightforward

interpretation. The first factor can be interpreted as trust in institutions, the second factor as trust in strangers, and the third factor as trust in known others. The second component consists of all the items of SOEP-trust and can thus be interpreted as "trust in strangers". This clearly shows that SOEP-trust measures the specific trust people have in strangers. It can clearly be distinguished from trust in institutions and known others.

Table 1: Dimensions of Trust

	Factor1 Trust in institutions	Factor2 Trust in strangers	Factor3 Trust in known others
How much trust do you have in...			
... parliament	**0.742**	0.173	-0.006
... public authorities	**0.715**	0.139	0.107
... the European Union	**0.686**	0.163	0.004
... courts	**0.665**	0.085	0.092
... large companies	**0.581**	-0.003	0.033
... churches	**0.460**	0.218	0.193
... schools and the educational system	**0.564**	0.088	0.193
... press	**0.550**	0.076	0.081
... labour unions	**0.493**	0.015	0.028
... police	**0.584**	-0.015	0.273
... your own family	0.070	-0.051	**0.647**
... neighbours	0.115	0.165	**0.716**
... friends	0.045	0.145	**0.695**
... strangers	0.183	**0.636**	0.091
To what extent do you agree or disagree?			
In general, you can trust people.	0.155	**0.647**	0.268
Nowadays, you can't rely on anybody.	-0.106	**-0.666**	-0.229
It's better to be cautious before trusting strangers.	-0.050	**-0.685**	0.154

Notes: Factor analysis using the principal components factor method and an orthogonal varimax rotation. The table reports the rotated factor loadings for the three factors with an eigenvalue larger than 1.

Source: AS02, AS03, AS04, AS05, and AS06 with a total of 3,180 observations.

For each of the three dimensions we calculated count indices[3]. The reliability of these scales measured by Cronbach's alpha is quite good. It is 0.82 for the index on trust in institutions, 0.62 for trust in known others, and 0.66 for SOEP-trust. Having introduced these three measures of trust we are interested how sensitive these measures are.

2.2 Sensitivity of SOEP-trust

In this section we assess the sensitivity of SOEP-trust. We check whether the position in the survey matters, whether there is a social desirability bias, and whether SOEP-trust is a stable and reliable measure.

2.2.1 Position in the Survey

We varied the position of the items of SOEP-trust in the AS06. Subjects were randomly divided into two groups. In one group, the trust questions were asked early in the questionnaire (number 33 out of 118 questions) and in the other group towards the end (question 93). We compare the ranking and the variance of this variable. The latter is clearly not dependent on the position in the survey (Levene's robust test statistic for the equality of variances: $F(1, 1039) = 0.169$ $P > 0.681$). However, the ranking is affected. Respondents who were asked the questions late in the survey exhibit more trust in strangers than people who answered the question early in the questionnaire (Table 2). Since the four preceding questions were the same in both situations, this effect is not likely to be driven by different preceding questions but by the position in the survey. Although the size of the effect is rather small, this suggests that to compare trust across time or space, the items have to have a similar position within the survey.

[3]The count indices are the mean answer of the non-missing items that load highest on each of the three dimensions. The bold numbers indicate which item loads highest on which dimension. The value for a person is calculated as soon as at least two items per component are non-missing. For the count index SOEP-trust we additionally made sure that at least one caution item and one trust item was included. For the count index trust in known others, we additionally included an item about trust in co-workers. This item was not included in the factor analysis, since only people with a job are asked. An inclusion would thus exclude a major share of the population.

Table 2: Position of the trust questions within the survey

To what extent do you agree or disagree? *(percentage agreeing)*	**Position in survey** **early**	**late**	**Wilcoxon rank-sum test (p)**
- In general, you can trust people.	72.8%	73.4%	> 0.298
- Nowadays, you can't rely on anybody.	36.3%	32.3%	< 0.073
- It's better to be cautious before trusting strangers.	89.4%	88.2%	< 0.023
Count-index: SOEP-trust	**early** $<$ **late**		< 0.017

Source: AS06 with a total of 1,033 observations.

2.2.2 Social Desirability

We measure social desirability by the Balanced Inventory of Desirable Responding (BIDR) developed by Paulhus (1991). Based on this, Winkler et al. (2006) developed a short version that is suitable for large surveys in the general population. The BIDR has two dimensions. One is called "self-deceptive enhancement" and captures a tendency to see reality in a more optimistic way than justified. This self-deception is not thought to be conscious. The other dimension is called "impression management" and measures the degree to which a person consciously tries to construct a favourable picture of other people. Since trust is desirable in society, trust questions are found to be correlated with scales of social desirability (Rotter 1967).

We find that the dimension "impression management" is significantly positively correlated with the survey measures SOEP-trust (Spearman's rank correlation: $\rho = 0.11$), trust in known others ($\rho = 0.15$) and trust in institutions ($\rho = 0.14$). The correlations of survey trust measures with "self-deceptive enhancement" are quite low and only significant for SOEP-trust ($\rho = 0.07$, $p < 0.051$) and trust in known others ($\rho = 0.08$, $p < 0.036$). The correlation is not significant with trust in institutions (ρ = -0.00).

Thus, people who are subject to a social desirability bias are likely to overstate their trust in strangers, known others, and institutions. This calls into question the validity of these measures of trust and makes comparisons across space and time difficult.

2.2.3 Stability and Reliability

We assessed the stability of SOEP-trust by repeating core questions of the AS05 six weeks later with a sub-sample (n = 193). If trust measured in a survey is a stable measure, we expect a correlation close to one. We find that SOEP-trust is only a moderately stable measure since the Spearman's rank correlation coefficient between the two time periods is 0.48[4]. The trust level did not change for one-third of the participants, increased for 37%, and decreased for 30%. Since trust is only moderately stable over time, the reliability of SOEP-trust cannot be assessed through a simple correlation. Instead we used a composite reliability test according to Raykov (2004), which tests whether a single factor underlies a certain set of variables. A structural equation model approach is used to calculate the reliability coefficient ρ. We find that it is 0.81. This shows that our measure of SOEP-trust is a reliable measure of trust. In sum, though SOEP-trust is only moderately stable across time, it does prove to be a reliable measure of trust.

2.3 Validity of SOEP-trust

In the previous section we showed that our new survey measure of trust has a high level of reliability but is sensitive to the position and also has a social desirability bias. In this section, we assess the validity of the new measure in two different ways. First, we used a survey measure of past trusting behaviour (Glaeser et al. 2000). Relying on self-reports, we asked people: "How often do you..."

- ...lend personal possessions to your friends (CDs, books, your car, bicycle etc.)?
- ...lend money to your friends?
- ...leave your door unlocked?

Respondents' answers were either "never", "infrequently", "sometimes", "often", or "very often". A factor analysis confirms the one-dimensionality of these three items.

[4]The items of SOEP-trust are correlated as follows: "In general you can trust people" with 0.41, "Nowadays, you can't be too careful" with 0.45 and "It's better to be cautious before trusting strangers" with 0.34.

We interpret this as past trusting behavior trusting . The reliability of these items in a count index is 0.56. A first indication that the new trust measure is valid stems from the fact that it correlates significantly with past trusting behavior (Table 3: Spearman's $\rho = 0.17$). Concerning the other two dimensions, we find that trust in known others correlates significantly with past trusting behaviour (Spearman's $\rho = 0.08$) whereas the correlation with trust in institutions is essentially zero and not significant. The latter is not surprising since the past trusting behaviour has little to do with trust in institutions. A second way of assessing the validity of the new trust measure is to compare it to other established measures of trust. Apart from the formulation in the World Values Survey, we used two other well-established measures of trust, namely the question from the European Social Survey (ESS) and the original 10 items of the dimension "agreeableness" from the NEO-PI-R that measures trust. The ESS-trust question simply asked people on an 11-point Likert scale how much trust they have in others. As can be seen in Table 3, this ESS-trust correlates highly ($\rho = 0.47$) with our new measure of trust. A similar correlation of 0.55 can be observed between the new measure of trust with the trust factor of the NEO-PI-R and of 0.53 with the GSS question.

From this we conclude that SOEP-trust is a valid measure of trust in strangers, and is well connected to existing survey measures of trust.

3 Using Experiments to Measure Trust

In this section we assess the sensitivity and validity of EXP-trust in a common framework. Despite its frequent use in economic research, we are not aware of any other study that analyses the sensitivity and validity of the trust game in depth and at the same time in the same framework. Furthermore, we have chosen to run the experiment with a representative sample in order to compare the experiment directly with representative surveys. Since not all randomly selected persons agree to participate in an experiment, it is important to check whether results from the experiment can be generalised to a whole population. That is why we start by analysing a possible selection effect, then

Table 3: Correlations between different concepts of trust in surveys: ESS, NEO-A1, and GSS

	self-report. trusting behaviour	**ESS**	**NEO-A1**	**GSS**
SOEP-trust	0.17***	0.47***	0.55***	0.53***
- In general you can trust people.	0.10***	0.37***	0.49***	0.47***
- Nowadays, you can't rely on anybody.	-0.15***	-0.34***	-0.40***	-0.39***
- It's better to be cautious before trusting strangers.	-0.13***	-0.19***	-0.31***	-0.48***
- Trust in first-time met stranger	0.08***	0.34***	0.33***	0.22**
Trust in known others	0.08***	0.29***	0.31***	0.34***
Trust in Institutions	0.02	0.28***	0.31***	0.29***

Notes: Spearman's rank correlations. The table reports the correlation coefficients and the significance level is denoted as follows: $*$: 10% $**$: 5% $***$: 1%

Source: AS02, AS03, AS04, AS05, and AS06 and the data on the GSS trust question was collected in a separate student survey at Royal Holloway, University of London in June 2008.

analyse its sensitivity and validity.

3.1 Design of the Experiment

The design of the experiment to measure trust is based on the investment game introduced by Berg et al. (1995). Two players anonymously interact with each other in the following way. The first mover gets an endowment of 10 points and can transfer zero to ten points to the second mover. Every point that is transferred is doubled by the experimenters. The second mover also gets an endowment of ten points. After receiving points from the first mover, he/she decides on how much of the endowment to transfer back to the first mover (zero to ten points). As with the first mover's transfer, the back-transfer by the second mover is doubled by the experimenters. After the second mover's decision, the game ends and the subjects are paid their income in euro (one point equals one euro) by cheque sent a few days later.

This design was developed by Fehr, Fischbacher, Schupp, von Rosenbladt & Wagner (2002) and it was implemented in the SOEP 2003 and in the AS02. In the other two accompanying studies, we used a small variation of the design. In the AS03, we

implemented the strategy method for the second mover. In order not to make it too complicated, we restricted the options of both players to transfer either zero, five, or ten points to the other player. In the AS04, we removed the option to transfer half of the endowment to the other player by allowing them to transfer only zero, two, four, six, eight, or ten points to the other player. In the SOEP 2003 we implemented a high-stakes treatment that used an exchange rate of one point to ten euros for a small part of the sample (117 out of 1,432). Thus, both players had an endowment of one hundred euros.

The Nash equilibrium of this trust game can be described as follows. If we assume that both players are rational and selfish and that this is common knowledge, neither one of the players ever transfers a single point to the other. We can relax the assumption that both players are selfish and instead assume that both players are inequality-averse as described by Fehr & Schmidt (1999). If the second mover is inequality-averse enough($\beta > \frac{1}{3}$), his/her back-transfer is equal to the first mover's transfer. If the second mover's inequality-aversion is not common knowledge, the first mover's transfer depends on his/her belief about the probability that the second mover will equalise the inequality. If this belief is above a certain threshold, the first mover transfers the whole endowment and otherwise nothing at all. One can show that the more inequality-averse the first mover, the higher this threshold is. Another relaxing assumption is that players have a preference for reciprocity instead for complete selfishness (Falk & Fischbacher 2006). With this assumption, the second mover's back-transfer increases weakly with the strength of his/her preference for reciprocity. Note that the prediction is not only a back-transfer of zero or equal to the first mover's transfer, but that transfers in between are possible as well. However, the first mover's transfer is predicted to be either zero points or the whole endowment. He/she will transfer zero points if his/her second-order beliefs are low enough, that is, below a certain threshold. Otherwise, first movers will transfer the whole endowment. This threshold increases with the first mover's degree of reciprocity. Thus, the more reciprocal a first mover is, the less he/she is predicted to transfer in a trust game.

In sum, a preference for equity or reciprocity increases the likelihood that the second

mover is behaving in a trustworthy manner. However, it decreases the probability that the first mover trusts given the second mover's preferences.

3.2 Selection

Imagine the situation that someone knocks on your door and asks you to participate in a survey on politics and society that would last about an hour. Included in the survey is a "game" in which you can earn some money. Many people will be suspicious and mistrust the person at the door. This situation is common for the interviewers in the social research section of TNS Infratest in Munich. They conduct the interviews for SOEP's accompanying studies as well as for the panel study itself. Besides other factors, the distrust in the interviewer and/or the survey organisation may lead people to refuse to participate in the survey.

If one wants to study trust on a representative level, it is therefore crucial to avoid a randomisation bias (Heckman & Smith 1995) due to trust. The issue has grown in importance in light of recent criticism of experimental economics (Levitt & List 2007), which is confronted with the same selection problem. So far, however, there has been little discussion about selection into either lab or field experiments. In one of the few studies that has addressed the problem, Bellemare & Kroeger (2007) do not find any randomisation bias for a trust experiment in a random sample of 541 regular panel participants in the Netherlands. Contrary, Harrison et al. (forthcoming) find a randomisation bias in a sample of 253 Danish subjects who were not part of a panel study but recruited for a "snapshot study" like our accompanying studies. They found that risk-averse people were less likely to participate in their study. These two studies indirectly assess the randomisation bias for the variable of interest.

The aim of this section is to examine the two possible randomisation biases in our study design. The first is that less trusting persons will be less likely to participate in the survey and second, that conditional on participation in the survey, it is less likely that they will agree to participate in the experiment at the end of the questionnaire. It is important to analyse these effects separately, since survey respondents know what the

survey is about, while participants in an experiment normally do not see its ultimate purpose.

The participation bias in the survey can be addressed by weighting subjects such that the distribution of basic variables conforms to the German population (Kroh & Spieß 2008). We can then compare the level of trust with and without weights. Different levels would then be an indication that selection is an issue. Indeed we do not find any indication that the weighted mean is significantly different from the unweighted one in the panel or in the accompanying studies. The average weighted transfer (mean: 5.15; n = 1454) lies in the 95% confidence interval of the unweighted mean (5.02 - 5.31). This holds also for the panel and the accompanying studies separately.

The second randomisation bias, refusal to participate in the experiment, can be addressed by comparing trust measured in the survey for participants in the experiment with non-participants. Since all the subjects who refused to participate in the experiment completed the survey, we are able to directly assess their trust through trust measured in the survey. As a proxy for trust in the interviewer and the survey organisation, we take SOEP-trust and past trusting behaviour. The latter captures the experience of past interactions that involved trusting others. We find that in the SOEP, past trusting behaviour nor SOEP-trust are related to the refusal to participate in the experiment (Mann Whitney test: past trusting behaviour $z=0.673$ $p>0.500$; SOEP-trust 0.505 $p>0.613$). Contrary to the SOEP, we find in the accompanying studies a significant lower level of trust among people who refused to participate in the experiment than those who participated (Mann Whitney test: past trusting behaviour $z=2.23$ $p<0.026$; SOEP-trust 1.322 $p>0.185$).

One likely explanation for a randomisation bias due to trust in the AS but not in the SOEP is based on the different set-ups. In the SOEP, subjects are familiar with the survey organisation and usually also with the interviewer, since these persons have been participants in the panel for three years. In accompanying studies, on the other hand, people are coming into contact with the survey organisation and the interviewer for the first time. In the accompanying studies, we find that in 2003, 5.1% and in 2004,

10.8% of subjects refused to participate in the experiment. In the panel study SOEP, only 4.8% of 1,504 subjects who completed the questionnaire refused to participate in the experiment in 2003. The difference in refusal rates between AS and SOEP is highly significant (Fisher's exact test: $p<0.001$).

This interpretation is supported by the following two facts. First, the null effect in the SOEP cannot be explained by a ceiling effect, since past trusting behaviour is even slightly higher in the accompanying studies than in the SOEP 2003. Second, we find that people who refuse to participate in the experiment are also more likely to leave the panel in the following three years (Fisher's exact test: $p<0.001$). However, the people who left the panel did not exhibit different trust levels in the experiment (Mann-Whitney test: $z=0.387$ $p>0.698$) or in the survey (Mann-Whitney test: past trusting behaviour $z=0.849$ $p>0.395$; SOEP-trust $z=1.529$ $p>0.126$).

In the accompanying studies, people were asked about their willingness to participate in a similar study another time. Similar to the panel study, we find that people who refuse to participate in the experiment are also less willing to participate another time in a similar study (Mann-Whitney test: $z=10.641$ $p<0.000$). Unlike in the panel study, we find that people who are less willing to participate in another similar study exhibit less trust in the experiment (Spearman's rank correlation $\rho = 0.08$ $p<0.034$) as well as in the survey (Spearman's rank correlation: past trusting behaviour $\rho = 0.13$ $p<0.001$; SOEP-trust $\rho = 0.02$ $p>0.504$).

Concerning risk-aversion we find very similar results as those for trust and found in Harrison et al. (forthcoming). We measured people's risk aversion by asking "Are you generally a person who is fully prepared to take risks, or do you try to avoid taking risks?". People answered on a Likert scale. This measure of risk-aversion is shown to be a valid measure by comparison to a lottery experiment with real monetary stakes (Dohmen et al. 2007). As with trust, we find that in the panel study SOEP, [5] there is no significant relation between risk preferences and the participation rate (Mann-Whitney

[5]The question on risk aversion was asked one year after the experiment in 2004. We assume that risk preferences are stable over this time period

test: z=1.487 p>0.136), whereas in the AS03 the most risk-averse people are less likely to participate in the experiment than the least risk-averse (Mann-Whitney test: z=2.135 p<0.033).

Thus, our results suggest that representative trust games are subject to selection effects due to trust and risk preferences if subjects are unfamiliar with the situation they are in (e.g., unknown survey organisation and/or interviewer).

In Appendix A.3, we additionally analyse other factors that might influence the participation in the experiment, including social preferences, personality, and demographic variables. In sum we find that selflessness, reciprocity, and interviewer characteristics do not matter in the decision to participate in the experiment. A medium income, a large household size, and living in Eastern Germany reduce the probability that a person will refuse to participate. Finally, the longer the previous questionnaire lasted, the more likely it is that participants will refuse to participate in the present experiment in the AS.

In sum, we found that the level of trust is related to the decision to participate in the experiment for subjects who are participating for the first time in this kind of interview. On the other hand, trust has no influence if subjects are familiar with the general set-up of the study. Based on these findings we conclude that a longitudinal panel survey where a "trust relationship" between the survey institute and the respondents is already established is the best way to minimise the total response error when adding experimental add-ons.

3.3 Validity of experimental measure

The decision to trust is influenced by people's preferences and expectations. Risk and social preferences and expectations about the other player's behaviour are expected to shape the decision to trust (Coleman 1990). Thus, one can check the validity of the trust measure by showing that these preferences and expectations indeed correlate with EXP-trust.

We measured people's risk-aversion as outlined in section 3.2. We find that the

first mover transfer is higher the less risk-averse people are in AS03 (Spearman's rank correlation $\rho = 0.13$ p<0.012).[6] Thus, people's risk preferences influence their decisions to trust or mistrust.

Selflessness is one component of social preferences, and we measured selflessness by asking people how often they volunteer in clubs and social services. They answered by indicating whether they volunteered "never", "seldom", "at least once a month", "at least once a week", or "daily". This question is asked in every AS and in the SOEP. We find that the more often people volunteer — thus behaving more selflessly — the more they transfer to the second mover (Spearman's rank correlation $\rho = 0.06$ p<0.016).

In the AS03 we elicited first movers' expectations by asking how much they expected the second mover to transfer back if they were to transfer zero, five, or ten points to the second mover. A selfish first mover would like to maximise his/her payoff. We therefore calculated which of the three transfers (zero, five, or ten) a first mover expected to maximise his/her payoff.[7] The higher the transfer needed in order maximise expected profits, the more a selfish first mover is expected to transfer. We find a positive and significant correlation (Spearman's rank correlation $\rho = 0.20$ p<0.003). However, a non-selfish first mover is not primarily interested in making the transfer that he/she expects to maximise profits. If he/she cares about inequality, he/she would like to know which of the three transfers yields the lowest inequality.[8] With our data, this is also possible to calculate, and the higher the transfer expected to minimise inequality is, the higher we expect the first movers' transfers to be. This other measure for expectations is positively correlated with first mover's transfer, as well (Spearman's rank correlation $\rho = 0.31$ p<0.001). As a third measure, we calculate the average expected back-transfer, which can be interpreted as a general measure for the expectation of the second mover's selflessness. Again, we find a positive correlation (Spearman's rank correlation is: $\rho =$ 0.18 p<0.005).

[6] In the SOEP the survey measure of risk was implemented in 2004, that is, one year after the experiment took place. We still find a significant relation between first-mover transfers and the risk measure in the survey (Spearman's rank correlation $\rho = 0.08$ p<0.048).

[7] Where the profit was the same for two or three transfer levels, we chose the lowest of them.

[8] In case the inequality is the same for two or three transfer levels, we chose the higher of them.

In sum, we have shown that EXP-trust is influenced by risk and social preferences and by expectations about the second mover's behaviour. Further, in Section 4, we demonstrate that EXP-trust is correlated with survey measures of trust. All these results are strong indications that the EXP-trust is a valid measure of trust.

3.4 Sensitivity of experimental measure

3.4.1 Students versus general population

A common critique of laboratory experiments is that students, the preferred subject pool in experiments, may behave systematically differently than non-students (Levitt & List 2007). In addition, there are a number of studies showing that students behave differently than other groups or that economics students are different from non-economics students (e.g., Fehr et al. 2006). A stronger test of the claim that students behave systematically differently is to compare their behaviour to a representative sample of the general population. Only a few studies have been designed this way, and the following results are found. In the U.S. state of Vermont, Carpenter et al. (2007) found in a field experiment that students donated 17 dollars less to charities than non-students, who donated 72 dollars out of 100. In a representative ultimatum game in Taiwan, no difference was found between students and non-students (Fu et al. 2007). Concerning discount rates, Harrison et al. (2002) found that in Denmark, students have a six percentage point higher discount rate than non-students. In a similar study in Denmark, students were found to be more risk-averse than non-students (Harrison et al. 2007). Since students are found to be more risk-averse and less pro-social, previous studies have suggested that students transfer fewer points in trust games than non-students. Indeed, this result was found by Bellemare & Kroeger (2007) for the Netherlands, where students transferred much less in a trust experiment than a representative population sample.

Among our German sample of 1,665 first movers in the trust game, we identify 47 as students. We find the opposite of all previous studies in that students transfer 61% whereas non-students transfer 50% of their endowment. Thus, students exhibit a 21% higher level of trust than non-students. This is difference is highly significant (regres-

Table 4: Dependent variable: Level of trust (first mover transfer)

	(1)	(2)	(3)	(4)
Dummy for being a student	0.967**	0.992**	0.773*	1.219**
	(0.440)	(0.434)	(0.449)	(0.570)
Dummy for having a university		0.592***	0.464**	
degree		(0.193)	(0.203)	
Age			-0.037	
			(0.023)	
Age^2			0.000	
			(0.000)	
Household income: <2500 Euro			-0.083	
(Base: <1500 Euro)			(0.196)	
Household income: <3500 Euro			-0.105	
			(0.218)	
Household income: <5000 Euro			0.404	
			(0.248)	
Household income: >5000 Euro			1.029***	
			(0.321)	
Risk aversion: medium				0.309*
(Base: high)				(0.181)
Risk aversion: low				0.655***
				(0.246)
Constant	5.081***	4.986***	5.855***	4.881***
	(0.072)	(0.078)	(0.528)	(0.136)
N	1539	1539	1539	1056
Adjusted R^2	0.00	0.01	0.02	0.01

Notes: Models 1 - 4 are estimated using an OLS specification. Numbers in parentheses denote the robust standard errors. The number of observations is lower in model four since data on the survey question on risk aversion is only available for people in the SOEP and AS03. Significance levels are denoted : $*$: 10% $**$: 5% $***$: 1% *Source:* SOEP 2003 and AS02, AS03, and AS04

sion 1 in Table 4). The results from the survey measure of trust strongly support these findings. SOEP-trust, trust in institutions, and past trusting behaviour are more pronounced among students (Mann-Whitney test: SOEP-trust $z = 10.521$ $p<0.001$; trust in institutions $z = 2.88$ $p<0.005$; trust in known others $z = 0.01$ $p>0.99$; past trusting behaviour $z = 19.78$ $p<0.001$). Students are typically younger than the average population, have a lower income, and a higher level of education. Do these different observable characteristics explain the observed differences? Controlling for these characteristics, we find that students do not trust more because they have a higher education, although people with a university degree have a higher level of trust than those without. The difference between students and non-students is still highly significant and similar in magnitude to findings when not controlling for university degree (regression 2 in Table 4). Further, we controlled for age and income and found that these variables decreased the coefficient for the dummy for students by about 22% but remained weakly significantly different from zero (regression 3 in Table 4). Thus, in the German population with our survey measure of risk aversion, we find that students seem to be less risk-averse than non-students. Therefore it is important to control for whether lower risk aversion can explain the higher level of trust among students than non-students. As expected, risk aversion is a determinant of trusting behaviour but the differences between students and non-students remain highly significant (regression 4 in Table 4). Trustworthiness does not differ between students and non-students (regression 1 in Table 5). Thus, students exhibit a higher level of trust than non-students. Differences in age and income reduce the difference by 22% but different degrees of risk aversion cannot explain the fact that students are more trusting than non-students.

3.4.2 Stakes

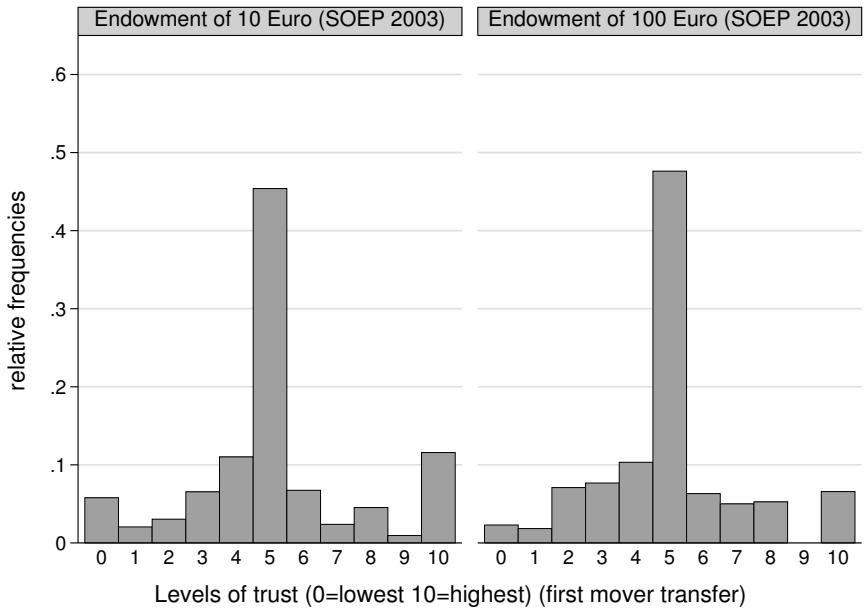
Only a few studies have examined the effects of stake size on behaviour in experiments. Most of these studies have analysed the ultimatum game and found that first movers' offers are independent of the stake size[9]. However, it is also found that the respondents'

[9] An exception was reported by Fu et al. (2007), who found that offers decrease with stakes.

minimal acceptable offer is lower with high stakes than with low stakes (Hoffman et al. 1996, Slonim & Roth 1998, Cameron 1999, Munier & Zaharia 2002, Fu et al. 2007). No significant effects are found in the dictator game (Forsythe et al. 1994, Cherry et al. 2002, Carpenter et al. 2005, List & Cherry 2008) or in the gift-exchange game (Fehr, Fischbacher & Tougareva 2002). Mutual trust measured in the centipede game is reduced significantly with lower stakes (Parco et al. 2002).

Johansson-Stenman et al. (2005a) are the only ones that have studied the effect of stake size in the trust game. In contrast to the findings in the gift-exchange game and the ultimatum game and in line with the centipede game, they found that first movers' transfers in rural Bangladesh were lower the higher the stakes. The proportions transferred to the second movers was 55% in the case with low stakes. The proportion transferred decreased to 46% in the medium-stakes condition, and to 38% in the high-stakes condition. The stakes were equivalent to 67, 337 and 1683 U.S. dollars. Concerning the behaviour of the second-mover, no difference was found for the different stake sizes.

Figure 1: Levels of Trust: Different stake sizes



We compared the differences between endowments of 10 and 100 euros. Contrary to

the results of Johansson-Stenman et al. (2005a), we find no differences in the average level of trust (Figure 1: t-test: t = 0.51 p>0.607). The high-stakes endowment was ten times higher than the low-stakes one, and so was the average transfer (5.16 versus 49.9 euros). The distributions also do not differ (Kolmogorov-Smirnov test: p>0.968). The levels of trustworthiness are not different in the two treatments either (regression 2 in Table 5). Our findings are based on stakes that are rather low compared to the average income in Germany rather low, but cover the size of many everyday trust relations.

3.4.3 Social desirability

Contrary to the questionnaire, the decision in the experiment remains private and is not communicated to the interviewer. Thus social desirability is not expected to influence the decision. We again use the Balanced Inventory of Desirable Responding by Paulhus (1991). We indeed find that impression management (Spearman's rank correlation $\rho =$ 0.08 p>0.145) and self-deception (Spearman's rank correlation $\rho = 0.03$ p>0.590) are not significantly correlated with EXP-trust. The decision to behave trustworthily or not is not correlated with social desirability either (regression 3 in Table 5).

Table 5: Sensitivity of experimental measure of trustworthiness

Dependant variable: dummy of being trustworthy	Students	Stakes	Social desirability
Dummy for being a student	-0.001		
	(0.048)		
Dummy for 100 Euro treatment		-0.072	
		(0.052)	
Soc. desir.: Impression (std.)			-0.020
			(0.028)
Soc. desir.: self-deception (std.)			-0.001
			(0.028)
Controlled for first mover transfer	YES	YES	YES
Constant	0.535***	0.518***	0.606***
	(0.020)	(0.047)	(0.058)
N	1912	602	292
Adjusted R^2	0.131	0.166	0.014
Cluster on individual level	YES	NO	NO

Notes: OLS regression with robust standard errors : $*$: 10% $**$: 5% $***$: 1%

Source: AS02, AS03, AS04 and SOEP 2003

4 Comparing Experimental and Survey Measures of Trust

It yet remains uncertain if our experimental measure captures the same kind of trust as is described by the various survey measures of trust. Previous research yielded ambiguous answers when comparing EXP-trust to GSS-trust. Given the criticisms lodged against this measure, this may not come as a surprise. In GSS-trust, it is unclear, for example, what or who it is that we trust. In Section 2.3 we showed that GSS-trust is correlated with several dimensions of trust such as trust in strangers, trust in known others, and trust in institutions. Unclear relations between GSS-trust and the experimental measure of trust are probably due to weakness of the survey question rather than problems with the experiment. Further indications for this can be found in Glaeser et al. (2000) and Gachter et al. (2004). Both studies report that their experimental measure of trust is not correlated with GSS-trust.

However, Glaeser et al. (2000) and Gachter et al. (2004) find that their experimental measures of trust are clearly correlated with survey questions on past trusting behaviour and a question on trust in strangers, which is formulated as the statement "You can't count on strangers anymore" and is similar to statements on which SOEP-trust is built. Thus, the question should not be whether survey measures of trust are correlated with experimental measures or not, but rather what kind of trust the trust game measures. To this effect, we measured trust in strangers (SOEP-trust), trust in known others, trust in institutions, and past trusting behaviour.

With our surveys, we have all the ingredients needed to test what kind of trust the trust game (EXP-trust) actually measures. We find that EXP-trust is significantly correlated ($\rho = 0.12$) with SOEP-trust (Table 6). Not only the overall measure (SOEP-trust) but also all its components are correlated with the experimental measure. Furthermore, EXP-trust is not correlated with the index "trust in institutions" nor the index "trust in known others". The fact that none of the single items of these two indexes are correlated significantly with EXP-trust further confirms that the experimental measure specifically measures trust in strangers.

Table 6: Correlations of different survey measures of trust with EXP-trust

	Spearman's ρ	sign.-level	obs.
Different indices of survey measures of trust			
SOEP-trust	0.116	0.000	1661
Trust in institutions	0.022	0.493	952
Trust in known others	0.013	0.682	949
past trusting behaviour	0.156	0.000	1654
To what extent do you agree or disagree?			
In general you can trust people.	0.066	0.007	1660
Nowadays, you can't rely on anybody.	-0.107	0.000	1702
It's better to be cautious before trusting strangers.	-0.099	0.000	1704
How much trust do you have in...			
strangers	0.137	0.000	732
your own family	-0.008	0.814	946
neighbours	0.010	0.768	947
friends	0.038	0.239	946
co-worker	0.041	0.257	764
churches	0.001	0.965	940
schools and the educational system	0.054	0.102	930
press	-0.014	0.669	945
labour unions	-0.013	0.688	923
police	0.027	0.411	948
parliament	0.034	0.292	943
public authorities	0.025	0.439	945
the European Union	0.018	0.591	923
courts	0.037	0.259	941
large companies	-0.034	0.301	931
How often do you ...			
lend personal possessions to friends?	0.140	0.000	1653
lend money to your friends?	0.097	0.000	1653
leave your door unlocked?	0.107	0.000	1692
Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair? (GSS-fair)	0.067	0.006	1679
Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves (GSS-help)	-0.004	0.880	1683

Source: SOEP 2003 and AS02, AS03, and AS04

Furthermore, we find that self-reported past trusting behaviour is significantly correlated with the experimental measure of trust ($\rho = 0.16$) and that all the three items of the index 'past trusting behaviour', – lending money, lending possessions, and leaving the door unlocked – are significantly related to experimental trust.

We additionally analysed two frequently used survey questions on fairness and helpfulness that are implemented in the GSS as well. The questions are *"Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair?* (GSS-fair) and *"Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves"* (GSS-help). We find that trusting behaviour is significantly positively correlated with fairness measured by GSS-fair but that the correlation is rather low ($\rho = 0.07$). GSS-help, on the other hand, is not related to trusting behaviour.

We have thus shown that trust measured in the experiment actually measures a specific kind of trust, namely trust in strangers. Whereas this in itself is perhaps not surprising, given the nature of the experiment where participants are in fact unknown strangers to each other, it is noteworthy on the other hand that EXP-trust has almost nothing to do with trust in institutions nor with trust in known others.

The correlations of trust measured in the experiment and SOEP-trust ($\rho = 0.12$) and past trusting behaviour ($\rho = 0.16$) are rather low, although they are significantly different from zero. The reason for this could be that the survey measure SOEP-trust mainly measures expectations of other people's trustworthiness. The statements "In general, you can trust people" and "Nowadays, you can't rely on anybody" give strong indications of how a person sees strangers. In the experiment, however, the decision to trust or not is not only influenced by expectations of others' trustworthiness but also by risk and social preferences (Naef et al. 2009). If the measure SOEP-trust in fact mainly measures the expectation part of the motivation to trust, then we would expect the correlations between the survey measure and expectations in the experiment to be rather high – in particular, higher than the correlations with the actual decision (for a related argument see also Sapienza et al. (2007)).

Table 7: Correlations between Expectations and behaviour in the experiment

	transfer expected to be ...		**average expected back-transfer**	**first-mover transfer**
	inequality minimising	**profit maximising**		
SOEP-trust	0.19***	0.17***	0.23***	0.12***
Trust in known others	0.09	0.08	0.11*	0.01
Trust in institutions	0.13**	0.12*	0.15**	0.02
past trusting behaviour	0.17***	0.15**	0.18***	0.16***
(number of observations)	(256)	(246)	(246)	(var.)

Notes: Spearman's rank correlations. The table reports the correlation coefficients and the significance level is denoted as follows: $*$: 10% $**$: 5% $***$: 1%

Source: AS03

We measured first movers' expectations of the behaviour of the second movers as outlined in Section 3.3. We calculate three measures that are: "the transfer expected to be profit-maximising", "the transfer expected to be inequality-minimising", and "the mean expected back-transfer". We indeed find that survey measures of trust correlate more strongly with expectations than with first mover transfers (Table 7). Interestingly, expectations correlate not only with SOEP-trust but also with our measure for trust in institutions as well. Thus, expectations seem to be less specific to one dimension of trust than is the first mover decision in the trust game. Whereas expectations measure trust in institutions and SOEP-trust, the decision in the experiment is only related to the survey measure for SOEP-trust.

Trust is a widely researched concept and many different factors have been found to be correlated with trust. It would be interesting to see if these factors are correlated with the experimental measure of trust similarly – or differently – than how they are correlated with survey measures of trust. As a further illumination of the relation of survey and experimental measures of trust we investigate exactly this question. In order to answer this question, we chose prominent factors that previous research has shown to be correlated with trust. As measures of trust, we compare trusting behaviour in the experiment, SOEP-trust, and the different measures of expectations in the experiment as explained above. Several studies have reported that socio-economic variables such as

Table 8: Different Trust Measures and recently found Correlations of Trust

	Experimental trust (same obs.)	**SOEP trust (same obs.)**	**SOEP trust (all obs.)**	**transfer expected to be ... inequal. minim.**	**profit maxim.**	**av. exp. back-transfer**
Age	-0.04*	-0.01	-0.01*	-0.15***	-0.13***	-0.08**
Being female	-0.00	0.01	-0.02***	-0.06*	-0.05	-0.05
Education	0.09***	0.11***	0.16***	0.15***	0.12***	0.10***
Household Income	0.08***	0.14***	0.15***	0.11***	0.12***	0.06
Being foreigner	-0.05*	-0.02	-0.04***	-0.01	-0.03	-0.03
Living in East Germany	-0.06**	-0.11***	-0.09***	0.02	-0.03	0.01
Religious	0.00	0.12***	0.11***	0.07*	0.05	0.11***
Being undenominational	0.01	-0.08***	-0.06***	0.06	0.02	0.05
Risk aversion	-0.13***	-0.13***	-0.14***	-0.03	0.04	0.09
Negative reciprocity	-0.06*	-0.08*	-0.10***	0.02	0.02	-0.03
Freq. of Volunteering	0.06**	0.12***	0.13***	0.05	0.02	0.06*
Being an Entrepreneur	0.06*	0.04***	0.04***	0.04	-0.02	0.02
Being a shareholder	0.07*	0.05	0.12***	0.12**	0.10**	0.05
Appr. number of obs.	1,660	1,660	25,500	800	800	800

Notes: Spearman's rank correlations. The table reports the correlation coefficients and the significance level is denoted as follows: $*$: 10% $**$: 5% $***$: 1%

Source: SOEP 2003, SOEP 2005, AS02, AS03, AS04, and AS05

age, gender, income, education, nationality, and place of living are correlated with trust (e.g., Alesina & La Ferrara 2002, Bellemare & Kroeger 2007, Rainer & Siedler forthcoming, Sutter & Kocher 2007). We find that the behavioural trust measure and SOEP-trust are correlated in similar ways with socio-economic variables[10] (Table 8). The only exception is gender where we find lower trust in women than in men when measured by survey, but no gender difference in trust when measured by experiment. Concerning religion, our results for Germany differ from previous studies in other countries (e.g., Guiso et al. 2003). We find that people with no religious affiliation exhibit lower SOEP-trust and people who are actively religious have higher SOEP-trust. Both effects cannot be confirmed with the behavioural measure of trust. Risk and social preferences (volunteering and negative reciprocity) are correlated significantly with both the survey and the experimental measure of trust with similar magnitudes. Finally, trust is found to be

[10]We do not find that the relation of trust and age is quadratic

higher among entrepreneurs and shareholders (Guiso et al. 2006, Guiso et al. 2008). In our data, we find that both EXP-trust and SOEP-trust are higher for entrepreneurs and shareholders. Concerning expectations, we find that they are related to age, education, income, religiosity, and being a shareholder, but not or only marginally to the other factors. In sum, we find that trust measured by the experiment has similar correlations with factors that have been reported, as trust measured by survey. This is another indication that both - EXP-trust as well as SOEP-trust - are valid measures of trust.

5 Conclusion

We have developed a compact survey measure of trust in strangers (SOEP-trust) that takes into account recent criticism of the widely used GSS/WVS question. We rigourously tested this measure and find that it is a valid and reliable measure of trust in strangers. However, one has to be careful in using SOEP-trust since it is a sensitive measure. This has implications for the use of the survey question in international comparisons and over time. This does not necessarily devalue our measure as compared to other survey trust measures, as these most likely exhibit the same sensitivities; however, we did not explore these.

We analysed an experimental measure of trust extensively and most importantly, always in the same setting. We showed that there may be a selection of more trusting people into the experiment if the individuals are participating in such a survey for the first time, whereas in the panel study, we do not find that selection is an issue. The experiment is quite insensitive to various changes. We find that stakes, social desirability, strategy space, and use of the strategy method do not affect the behaviour in the experiment in significant ways. However, we find that students, who are typical subjects of lab experiments, behave differently than non-students in that they trust strangers more than non-students. This finding is confirmed by the survey, where we find that SOEP-trust is higher among students than among non-students. Furthermore, we show that trusting behaviour is influenced by people's risk and social preferences as

well as their expectations.

In combination, we find that the experimental measure of trust is significantly correlated with SOEP-trust, which is specifically aimed to measure trust in strangers; but not with an index of trust in institutions and an index of trust in known others. Furthermore, experimental trust correlates with related factors similarly as SOEP-trust does. We conclude that the common experimental measure of trust is a valid measure, which captures a specific form of trust: trust in strangers.

References

Alesina, A. & La Ferrara, E. (2002). Who trusts others?, *Journal of Public Economics* **85**(2): 207–234.

Ashraf, N., Bohnet, I. & Piankov, N. (2003). Is trust a bad investment, *Working paper* .

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

Berg, J., Dickhaut, J. & McCabe, K. (1995). Trust, reciprocity, and social-history, *Games and Economic Behavior* **10**(1): 122–142.

Brandts, J. & Charness, G. (2000). Hot vs. cold: Sequential responses and preference stability in experimental games, *Experimental Economics* **2 3**: 227–38.

Brosig, J., Weimann, J. & Yang, C.-L. (2003). The hot versus cold effect in a simple bargaining experiment, *Experimental Economics* **6 1**: 75–90.

Cameron, L. A. (1999). Raising the stakes in the ultimatum game: Experimental evidence from indonesia, *Economic Inquiry* **37**(1): 47–59.

Carpenter, J., Connolly, C. & Myers, C. (2007). Altruistic behavior in a representative dictator experiment, *Working paper* .

Carpenter, J., Verhoogen, E. & Burks, S. (2005). The effect of stakes in distribution experiments, *Economics Letters* **86**(3): 393–398.

Cherry, T. L., Frykblom, P. & Shogren, J. F. (2002). Hardnose the dictator, *American Economic Review* **92**(4): 1218–1221.

Coleman, J. S. (1990). *Foundations of social theory*, Belknap Press of Harvard University Press, Cambridge, Mass.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J. & Wagner, G. G. (2007). Individual risk attitudes: New evidence from a large, representative, experimentally-validated survey, *Working paper* .

Ermisch, J., Gambetta, D., Laurie, H., Siedler, T. & Uhrig, S. N. (2007). Measuring people's trust, *Working paper* .

Falk, A. & Fischbacher, U. (2006). A theory of reciprocity, *Games and Economic Behavior* **54**(2): 293–315.

Fehr, E., Fischbacher, U., Schupp, J., von Rosenbladt, B. & Wagner, G. G. (2002). A nationwide laboratory examining trust and trustworthiness by integrating behavioural experiments into representative surveys, *Schmollers Jahrbuch* **122**: 519–542.

Fehr, E., Fischbacher, U. & Tougareva, E. (2002). Do high stakes and competition undermine fairness? evidence from russia, *Working paper* .

Fehr, E., Naef, M. & Schmidt, K. M. (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment, *American Economic Review* **96**(5): 1912–1917.

Fehr, E. & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation, *Quarterly Journal of Economics* **114**(3): 817–868.

Forsythe, R., Horowitz, J. L., Savin, N. E. & Sefton, M. (1994). Fairness in simple bargaining experiments, *Games and Economic Behavior* **6**(3): 347–369.

Fu, T.-T., Kong, W.-H. & Yang, C. (2007). Monetary stakes and socioeconomic characteristics in ultimatum games: An experiment with nation-wide representative subjects, *Working paper* .

Gabriel, O. W., Kunz, V., Rossteutscher, S. & van Deth, J. W. (2002). *Sozialkapital und Demokratie: Zivilgesellschaftliche Ressourcen im Vergleich*, Vol. Band 24 of *Schriftenreihe des Zentrums fuer angewandte Politikforschung*, WUV Universitaetsverlag, Wien.

Gachter, S., Herrmann, B. & Thoni, C. (2004). Trust, voluntary cooperation, and socioeconomic background: survey and experimental evidence, *Journal of Economic Behavior & Organization* **55**(4): 505–531.

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

Guiso, L., Sapienza, P. & Zingales, L. (2003). People's opium? religion and economic attitudes, *Journal of Monetary Economics* **50**(1): 225–282.

Guiso, L., Sapienza, P. & Zingales, L. (2006). Does culture affect economic outcomes?, *Journal of Economic Perspectives* **20**(2): 23–48.

Guiso, L., Sapienza, P. & Zingales, L. (2008). Trusting the stock market, *Journal of Finance* **63**(6): 2557–2600.

Guth, W., Huck, S. & Muller, W. (2001). The relevance of equal splits in ultimatum games, *Games and Economic Behavior* **37**(1): 161–169.

Haile, D., Sadrieh, A. & Verbon, H. A. A. (2008). Cross-racial envy and underinvestment in south african partnerships, *Cambridge Journal of Economics* .

Harrison, G. W., Lau, M. I. & Rutstroem, E. E. (forthcoming). Risk attitudes, randomization to treatment, and self-selection into experiments, *Journal of Economic Behavior & Organization* .

Harrison, G. W., Lau, M. I. & Rutstrom, E. E. (2007). Estimating risk attitudes in denmark: A field experiment, *Scandinavian Journal of Economics* **109**(2): 341–368.

Harrison, G. W., Lau, M. I. & Williams, M. B. (2002). Estimating individual discount rates in denmark: A field experiment, *American Economic Review* **92**(5): 1606–1617.

Heckman, J. J. & Smith, J. A. (1995). Assessing the case for social experiments, *Journal of Economic Perspectives* **9**(2): 85–110.

Hoffman, E., McCabe, K. A. & Smith, V. L. (1996). On expectations and the monetary stakes in ultimatum games, *International Journal of Game Theory* **25**(3): 289–301.

Hoffman, E., McCabe, K. A. & Smith, V. L. (1998). Behavioral foundations of reciprocity: Experimental economics and evolutionary psychology, *Economic Inquiry* **36**(3): 335–352.

Holm, H. & Nystedt, P. (2008). Trust in surveys and games - a methodological contribution on the influence of money and location, *Journal of Economic Psychology* **29**(4): 522–542.

Johansson-Stenman, O., Mahmud, M. & Martinsson, P. (2005a). Does stake size matter in trust games?, *Economics Letters* **88**(3): 365–369.

Johansson-Stenman, O., Mahmud, M. & Martinsson, P. (2005b). Trust, trust games and stated trust - evidence from a field experiment in bangladesh, *Working paper* .

Kroh, M. & Spieß, M. (2008). Documentation of sample sizes and panel attrition in the german socio economic panel (1984 until 2006), *DIW Data Documentation 27* .

Levitt, S. D. & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world?, *Journal of Economic Perspectives* **21 2**: 153–74.

List, J. A. & Cherry, T. L. (2008). Examining the role of fairness in high stakes allocation decisions, *Journal of Economic Behavior & Organization* **65**(1): 1–8.

Miller, A. S. & Mitamura, T. (2003). Are surveys on trust trustworthy?, *Social Psychology Quarterly* **66**(1): 62–70.

Munier, B. & Zaharia, C. (2002). High stakes and acceptance behavior in ultimatum bargaining: A contribution from an international experiment, *Theory and Decision* **53**(3): 187–207.

Naef, M., Fehr, E., Fischbacher, U., Schupp, J. & Wagner, G. G. (2009). Decomposing trust: Explaining national and ethnical trust differences, *Working paper* .

Parco, J. E., Rapoport, A. & Stein, W. E. (2002). Effects of financial incentives on the breakdown of mutual trust, *Psychological Science* **13**(3): 292–297.

Paulhus, D. L. (1991). Measurement and control of response bias, *in* J. P. Robinson, P. R. Shaver, W. L. S. & F. M. Andrews (eds), *Measures of personality and social psychological attitudes*, Academic Press, San Diego, Calif., pp. 17–59.

Rainer, H. & Siedler, T. (forthcoming). Does democracy foster trust?, *Journal of Comparative Economics* .

Raykov, T. (2004). Estimation of maximal reliability: A note on a covariance structure modelling approach, *British Journal of Mathematical & Statistical Psychology* **57**: 21–27.

Reeskens, T. & Hooghe, M. (2008). Cross-cultural measurement equivalence of generalized trust. evidence from the european social survey (2002 and 2004), *Social Indicators Research* **85**(3): 515–532.

Rotter, J. B. (1967). A new scale for measurement of interpersonal trust, *Journal of Personality* **35**(4): 651–665.

Sapienza, P., Toldra, A. & Zingales, L. (2007). Understanding trust, *Working paper* .

Schotter, A., Weigelt, K. & Wilson, C. (1994). A laboratory investigation of multiperson rationality and presentation effects, *Games and Economic Behavior* **6**(3): 445–468.

Slonim, R. & Roth, A. E. (1998). Learning in high stakes ultimatum games: An experiment in the slovak republic, *Econometrica* **66**(3): 569–596.

Sutter, M. & Kocher, M. G. (2007). Trust and trustworthiness across different age groups, *Games and Economic Behavior* **59**(2): 364–382.

Vyrastekova, J. & Garikipati, S. (2005). Beliefs and trust: an experiment, *Working paper* .

Wagner, G. G., Frick, J. R. & Schupp, J. (2007). The german socio-economic panel study (soep) - scope, envolution and enhancements., *Journal of Applied Social Science Studies* **127**(1): 139–169.

Winkler, N., Kroh, M. & Spiess, M. (2006). Entwicklung einer deutschen kurzskala zur zweidimensionalen messung von sozialer erwuenschtheit, *Working paper* .

Yamagishi, T., Kikuchi, M. & Kosugi, M. (1999). Trust, gullibility, and social intelligence, *Asian Journal of Social Psychology* **2**(1): 145–161.

A Appendix

A.1 Overview of the different studies

Table A.1: Overview of the different studies

Study	No. of obs.		stake	scale	strategy	survey
	Survey	Exp.	size		method	sensitivity
AS02	442	442	10	11 options	no	
SOEP 03	22'611	1'432	10/100	11 options	no	
AS03	846	803	10	3 options	yes	
AS04	772	689	10	6 options	no	NEO-FFI
AS05	1'012					reliability
AS06	1'063					question order
						ESS question

A.2 Selection in the survey

Did only highly trusting people participate in the survey? If this were the case, we could not claim that our results are representative for Germany. We address this potential problem in three different ways. First we compare the level of trust when people are weighted such that the distribution in some basic characteristics conforms with the population distribution and when people are not weighted. Second, we check whether it makes a difference whether people are familiar with the survey organisation and/or the interviewer or whether the situation is unfamiliar by comparing the AS03 with the SOEP in 2003. Finally we know who has left the panel since 2003 and we can check whether this decision is dependent on their trust level in 2003.

In the accompanying studies, the unweighted mean of SOEP-trust (2.38) lies in the 95% confidence interval of the weighted mean (2.35 - 2.39). Similarly, no differences between the weighted and unweighted means are found for "trust in known others" or for "trust in institutions". In the SOEP, the unweighted mean of SOEP-trust (2.308) lies inside the 95% confidence interval of its weighted mean (2.29 - 2.31) as well. Further the weighted means in the AS03 (2.30) and in the SOEP 2003 (2.30) are not significantly different from each other (t-test: $t = 0.03$, $p > 0.974$). That is, it makes no differences for

SOEP-trust whether it is the first time people are interviewed, as in the accompanying study, or whether people have been in a panel for at least four years. The final test for a randomisation bias is that we know who left the panel between 2004 and 2006 and we know what their level of trust was in 2003. There is no difference in the mean of SOEP-trust between those who left the panel and those who stayed the following two years (t-test = 1.24 P > 0.21). In the accompanying studies we do not have this measure, but people were asked whether they would like to participate in similar study again. Again this can be used as an indication whether there is a randomisation bias due to trust. Again, we do not find any significant difference (t-test = 1.34 P > 0.17).

In sum we find that the participation in the survey in general is not influenced by how much trust people have.

A.3 Selection in the experiment

Beside trust we are able to test whether social preferences, personality characteristics, demographic variables, interviewer characteristics and the length of the questionnaire are determinants of refusal to participate in the experiment (Table A.2). As a proxy for social preferences we take the frequency with which subjects volunteer and participate in politics and citizens' initiatives. Again, we find that in the AS and in the SOEP 2003 there is no significant impact on the rate of refusal. Concerning the personality measures we find that positive and negative reciprocity do not predict refusal in the experiment either. However, some demographic variables do explain refusal to participate. There is a slight tendency for married people to be less likely to participate in the experiment. Further we find that in the SOEP 2003, people with a high or low household income are more likely to refuse to participate. In the AS, people in larger households are more cooperative in participating, whereas East Germans are less likely to participate. A further test is the length of the survey as a proxy for an additional response burden. The interviews in the AS04 were conducted using a laptop that recorded the time used for each question. We thus test whether the length of time from the beginning of the survey to the decision to participate in the experiment predicts this decision. We indeed

Table A.2: Dep = Dummy of whether a person refused to participate in the experiment

	SOEP03	AS03	AS04
Dummy of volunteering at least sometimes	-0.008	0.008	0.044
	(0.013)	(0.014)	(0.028)
Dummy of participating in political parties,	-0.005	0.039	-0.010
citizens' initiative	(0.017)	(0.029)	(0.027)
Negative reciprocity			0.009
			(0.007)
Positive reciprocity			0.015
			(0.011)
Dummy of being female	0.001	0.016	-0.028
	(0.011)	(0.010)	(0.022)
Age	0.000	0.000	0.000
	(0.000)	(0.000)	(0.001)
Intermediate secondary school	0.015	-0.012	-0.007
(Base: less than interm. secon. school)	(0.015)	(0.011)	(0.025)
High school and more	-0.007	0.004	0.017
	(0.014)	(0.015)	(0.035)
Living with partner (Base: Married)	0.010	-0.010	-0.049*
	(0.025)	(0.013)	(0.028)
Single or not living with partner	-0.009	-0.024*	-0.022
	(0.013)	(0.012)	(0.027)
Household income: <2500 Euro	-0.034***	0.003	-0.019
(Base: <1500)	(0.012)	(0.014)	(0.028)
Household income: <3500 Euro	-0.028**	0.015	-0.033
	(0.013)	(0.022)	(0.029)
Household income: <5000 Euro	-0.019	0.012	-0.042
	(0.015)	(0.027)	(0.032)
Household income: >5000 Euro	-0.009	0.008	-0.003
	(0.020)	(0.030)	(0.055)
Household size	0.007	-0.026***	-0.011
	(0.005)	(0.007)	(0.012)
Dummy of living in east Germany	-0.011	-0.012	-0.053**
	(0.012)	(0.011)	(0.021)
Dummy of being foreigner	-0.013		0.005
	(0.021)		(0.061)
Length of interview (in minutes)			0.002***
			(0.001)
Pr(refusal)	0.045	0.022	0.081
N	1480	716	659
log-likelihood	-278.27	-107.32	-193.79
Prob $> \chi^2_{crit.}$	0.581	0.014	0.062

Notes: Model 1 - 3 are estimated using a probit specification and the table reports the marginal effects of the different variables on refusing to participate in the experiment. Numbers in parentheses denote the standard error of the marginal effects. Significance levels are denoted : $*$ 10% $**$ 5% $***$ 1%
Source: SOEP 2003, AS03, and AS04

find that the longer the questionnaire, the less likely it is that participants agree to participate in the subsequent experiment in AS04.

Furthermore, the interviewer is a possible source of influence on participation in the experiment. However, we do not find interviewer characteristics such as age, gender and years of experience to be influential (models 1 - 3 in Table A.3).

Table A.3: Dep = Dummy of whether a person refused to participate in the experiment

	SOEP03	AS03	AS04
Years of experience in polling firm	0.001	-0.000	-0.002
	(0.001)	(0.001)	(0.003)
Dummy for a female interviewer	0.022	-0.024	-0.040
	(0.014)	(0.015)	(0.044)
Age of the interviewer	-0.000	-0.001	-0.002
	(0.001)	(0.001)	(0.002)
N	1491	729	558
Adjusted R^2	0.00	-0.00	0.00
Prob $> F_{crit.}$	0.116	0.229	0.559

Notes: Models 1 - 3 are estimated using a OLS specification and the table reports the coefficients. Numbers in parentheses denote the standard error clustered on interviewer level. Significance levels are denoted: $*$ 10% $**$ 5% $***$ 1%

Source: SOEP 2003, AS03, and AS04

In sum we find that selflessness, reciprocity, and interviewer characteristics do not matter in the decision to participate in the experiment. A medium income, a large household, and living in Eastern Germany reduce the probability that a persons will refuse to participate. Finally the longer the previous questionnaire lasted, the more likely it is that participants will refuse to participate in the present experiment in the AS.

A.4 Sensitivity of the experimental design

A.4.1 Design

In the AS02 and the SOEP 2003 we observed that the modus of the distribution of first movers' choices is to transfer half of the endowment (36% and 45%). This result is not an artefact of our specific study design or our instructions since in other studies

half of the endowment is the modal choice as well (e.g., Berg et al. 1995, Bellemare & Kroeger 2007). The question arises whether the widely observed pattern is dependent on the design of the trust game. More specifically we ask how the level of trust changes if we reduce the number of choices and remove the possibility to transfer half of the endowment. To test this, we run three different experiments. In the SOEP 2003, we have run the basic experiment with 11 options for a transfer, which run from 0 to 10 points. In the AS03 we reduced the choices to only three transfer options, which were 0, 5, or 10 points. In the AS04 we eliminated the choice for a transfer of 5 points by allowing only transfer levels of even numbers.

The distributions of transfers in the three experiments are very different by construction (Figure A.1). However, the average transfer was almost the same in the three experiments. In the SOEP 2003 with 11 options 51.5% of the endowments was transferred, in the AS03 51.3% and in the AS04 50.3% of the endowment was transferred. These differences are far from being significant.[11] The probability of behaving trustworthy is not dependent on how many options subjects have either (Figure A.2 and regression 1 in Table A.4).

A.4.2 Strategy method

The strategy method is a widely used elicitation procedure in experimental economics. With this method, second movers are asked to decide for every possible first mover decision. In the trust game, this procedure allows us to distinguish between selfish players and conditionally cooperative players and between the latter and altruistic players. If a second mover, for example, receives zero points from the first mover and he/she does not transfer back, we do not know whether he/she would transfer a positive amount back if a first mover transferred 5 points. The disadvantage is that it is more complicated to explain to subjects and the incentives are diluted since only one decision will actually be paid out. A further potential disadvantage is that the conditional decisions

[11] Two-sided T-test for differences in the mean with unequal variances: SOEP 03 vs. AS03 t = 0.69 p>0.48; SOEP 03 vs AS04 t = 0.60 p>0.54; AS03 vs. AS04 t = 0.06 p>0.95

Figure A.1: Levels of trust: different scales

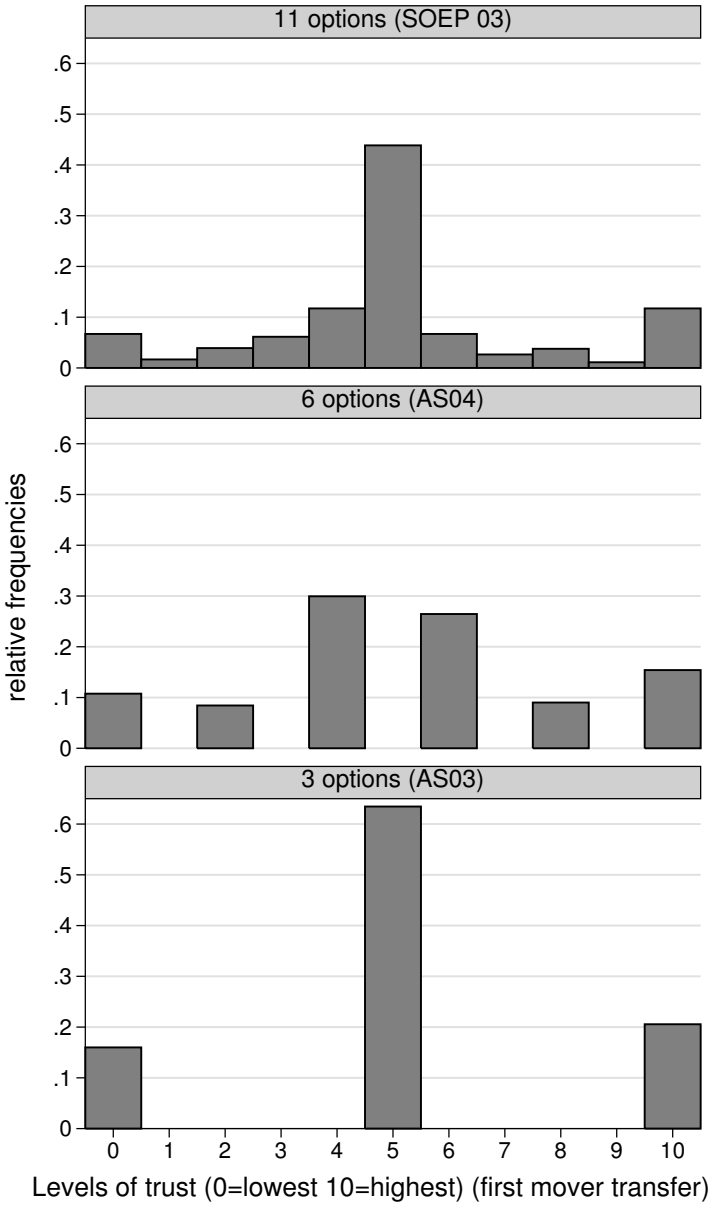
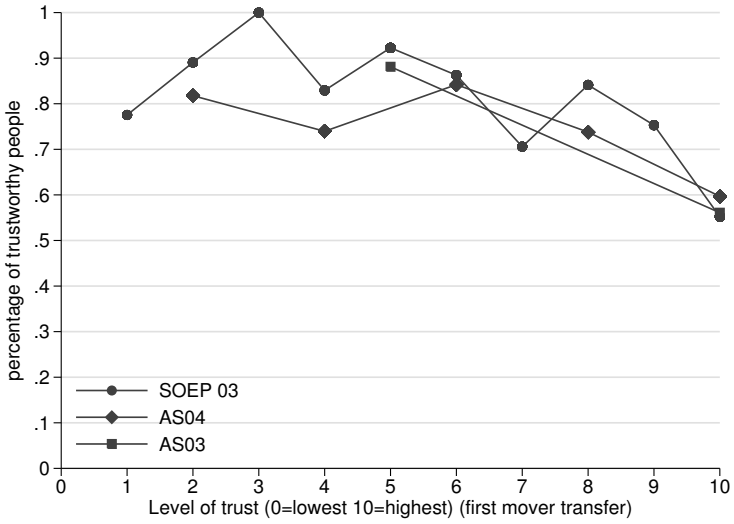


Figure A.2: Levels of trustworthiness



are less emotionally arousing than in the situation in which one knows what the other person decided. To our knowledge no study compares the strategy method and the direct method in a trust game. Previous research on bargaining games has shown that subjects' behaviour is different with the strategy method than with the direct method (e.g., Schotter et al. 1994, Hoffman et al. 1998, Guth et al. 2001, Brosig et al. 2003). Concerning social dilemma games, the only study that has analysed possible differences so far is a study by Brandts & Charness (2000). They analyzed a prisoner's dilemma and a game of chicken and found no difference. We implemented the strategy method for the second mover in the AS03. In order not to make it too complicated, we restricted first movers' choices to three options, which were 0, 5, or 10 points. In the other AS and the SOEP 2003 we used the direct method. Thus, we can compare the level of trustworthiness with and without the strategy method. We find no difference in the average level of trustworthiness between the strategy method and the direct method (regression 2 in Table A.4). For a first mover transfer of 5 (10) points, 91 (50)% of subjects are trustworthy using the direct method and 88 (56)% using the strategy method[12]. If we

[12]Two-sided Fisher Exact tests: 5 points $p>0.209$; 10 points $p>0.186$

analyse the decision of the second mover in more detail, we find a difference in average return transfers for a first mover transfer of 10 points. With the strategy method, return transfers are higher than with the direct method (6.0 versus 6.8 points: t-test: t=-2.603 p<0.01). For the other first mover transfers, that are 0 and 5 points, we do not find any significant difference in return transfers.

Table A.4: Sensitivity of experimental measure of trustworthiness

Dependent variable: dummy of being trustworthy	Design	strategy method
Dummy for design with 3 options	-0.009	
	(0.041)	
Dummy for design with 6 options	0.027	
	(0.025)	
Dummy for strategy method (AS03)		0.010
		(0.024)
Controlled for first mover transfer	YES	YES
Constant	0.551***	0.530***
	(0.023)	(0.024)
N	1722	1333
Adjusted R^2	0.126	0.160
Cluster on individual level	YES	YES

Notes: OLS regression with robust standard errors : $*$: 10% $**$: 5% $***$: 1%

Source: AS02, AS03, AS04, and SOEP 2003