Measuring reciprocity: Do survey and experimental data correlate?

Sandra Maximiano*

Purdue University
Department of Economics
Krannert School of Management
403 W. State Street, 47907 IN

January, 2012

Abstract

*I acknowledge Randolph Sloof for his valuable comments and participants at ESA conferences where gift exchange papers were presented and encourage the pursuit of the analysis presented this paper
1 Introduction

A large body of experimental evidence reveals that people are very heterogeneous in their preferences. Rather than being solely motivated by self-interest, individuals have concerns for fairness, others’ well-being, and behave reciprocal, i.e. they are prepared to sacrifice some money to reward (punish) good (bad) intentional behavior of others. In an overview article, Fehr and Gächter (2000) report that 40-66 percent of subjects display non-selfish behavior. In the context of incomplete contracts it has been shown that reciprocity can serve as a contract enforcement device. For instance, in a labor market context Fehr and Falk (1997) show that reciprocity gives incentives for selfish workers to provide a high effort if they expect a sufficiently generous wage offer. Other studies (Sonnemans and Sloof, 2001; Sloof and Sonnemans, 2003; Oosterbeek et al., 2006) show that reciprocity can act as an informal mechanism to alleviate the hold up problem, and thus increase the investment in firm-specific assets.

Given the established importance of reciprocity in the laboratory, the next step is to investigate whether the experimental findings persist in the real world. For that purpose we could either conduct field experiments, or use non-experimental data. In any case we need to measure reciprocity. In case of field experiments (and similarly in laboratory experiments) reciprocity is inferred from actually observed behavior. In general, in an experimental setting the most often used games to measure reciprocity are the gift exchange game (Fehr et al., 1993), and the trust (investment) game (Berg and McCabe, 1995). The first game mimics an incomplete (labor) contract environment. Here the employer moves first by offering the worker a contract. This contract usually consists of a wage and a required but non-enforceable effort level. Then, the worker decides whether to accept or reject the contract and in case the contract is accepted he or she chooses a costly effort level. The wage-effort correlation for individual...

1Alternatively, Levine (1998) give a type-based concept of reciprocity. Here reciprocity refers to the motivation that individuals have to reward a “good” person (i.e, someone with altruistic preferences) and punish a “bad” person (someone with spiteful preferences).
workers can be used as a measure of the worker’s willingness to reciprocate (positive reciprocity). In the trust (investment) game a trustor can send any amount $x$ of his initially endowment $y$ to a trustee. The experimenter triples $x$ on the way. Of the overall amount received, the trustee decides how much to send back, $z \in [0, 3x]$. The level of trust is measured by the amount $x$, and the level of trustworthiness is measured by $z/3x$. Trust can be interpreted as the anticipation of (positive) reciprocity and trustworthiness as the actual level of (positive) reciprocity. The trust and gift exchange game are suitable to measure positive reciprocity. Negative reciprocity is better measured by means of a moonlighting game (Abbink et al., 2000) where one player can take money from other, who can punish in return.

In case we use non-experimental two approaches are possible to measure reciprocity: (1) specify and estimate structural models that identify reciprocity parameters; (2) use subjective survey questions.

Surveys can be used to measure reciprocity, asking respondents for example about their general attitudes or past behavior. In fact, surveys are a classical method for gathering subjective data. They have been used by sociologists, psychologists and political scientists to measure respondents’ well-being, their feelings, beliefs, expectations and behavioral intentions. More recently, economists recognize that subjective questions may uncover interesting and important variables, but in general they remain sceptical regarding the quality of these data. Various reasons can explain why surveys may fail to elicit trustful responses, especially regarding sensitive issues.² Cognitive problems, such as miscomprehension, conceptual confusion, different understanding of the same questions, sensitiveness to the ordering of the questions, and imperfect recall of relevant information, account for some of the measurement error. Trustful answers may also be compromised due to certain biases. For instance, respondents may give a socially desirable answer rather than report what they actually

² Bertrand and Mullainathan (1991) summarize the literature and analyze the implications of using mismeasured subjective questions for empirical research.
do, think or believe in order to impress the interviewer or the researcher, or simply to preserve his or her self-esteem (social desirability bias). Moreover, people may respond differently to hypothetical situations compared to real situations (hypothetical bias) or they may respond as if they were another person they wished to be (idealized persona bias).

Experimental data is not totally immune to the biases referred to above. However, compared to surveys, experiments have the advantage of relying on actually observed behavior elicited in a controlled environment (laboratory experiments) and in an incentive compatible way. That is, it takes the hypothetical bias away. So, given that experiments offer us a more accurate measure of reciprocity, a question of interest then becomes, whether self-reported measures are significantly correlated with actual behavior. Our study aims at answering this question.

To our knowledge, the existing literature does not give an answer to this question. Only few studies relate survey-based measures to actual behavior and the focus is on trusting behavior in the trust game. Results are rather mixed. No survey questions have been identified that display an unambiguously high correlation with observed reciprocal behavior. In this paper we try to fill this gap. In particular, we analyze the data of attitudinal questions that subjects answered after participating in a bilateral gift exchange experiment. The questionnaire included some questions explicitly designed to measure reciprocity. Our main interest is to check whether answers to some questions correlate with workers’ reciprocal attitude. However, we also paid attention to employers’ behavior.

We find that questions that ask subjects directly about their concerns for reciprocity appear to be good predictors of workers’ actual propensity to reciprocate, as measured by the slope of the wage-effort schedule chosen. However, and in contrast to other studies, questions about trusting attitudes and past trusting behavior do not predict the workers’ choices made in the experiment. Those trusting questions instead
appear to predict employers’ wage offers.

This paper proceeds as follows. In the next section we briefly review related literature. Section 3 presents our experimental design and the survey questions, and offers some descriptive statistics. Section 4 presents the results and Section 5 concludes.

2 Related literature

Recently, some attention has been paid to the important research question of whether the answers to commonly used attitudinal survey questions are significantly correlated with actual behavior. However, the main focus has been on the measurement of trust and trustworthiness and contradictory results have been obtained concerning the predictive power of trust-related questions in explaining observed trust and trustworthiness.

The first study of interest is Glaeser and Souter (2000). They conducted a 137-question survey among 258 Harvard undergraduates, containing, apart from background questions, attitudinal and self-reported measures of trustfulness (questions about trusting attitudes and past trusting behavior) and trustworthiness. After three to four weeks, a subset of 196 subjects played two experimental games. The first one is the trust game as described in Section ?? with the investment doubled instead of tripled.3 The “second” game consists of subjects’ valuations for a series of envelops (containing 10 dollars) dropped at a public place. The envelops are all addressed to the subject, but they differ in some dimensions (e.g. sealed and stamped). Trust here is measured by the value on the dropped envelopes. They find that questions about trusting attitudes did not predict actual trusting behavior, while questions about past trusting behavior did. Also, questions related to trustworthiness do not predict trustworthiness behavior. However, the index of trust-related questions is significantly

3Given the survey stage, the anonymity of subjects is not guaranteed in Glaeser and Souter (2000), which increases cooperation and creates less variation in the data. The authors thus decided to double the amount $x$ rather than triple it. Even though, 71% of trustors sent the full amount $x$ to trustees.
correlated with the experimental measure of trustworthiness. Therefore, the authors conclude that “...the best way to determine whether or not a person is trustworthy is to ask him whether or not he trusts others.” (Glaeser and Souter, 2000, p. 833)

Karlan (2005) reaches the same conclusion. He considers a group of overall 864 participants in a microcredit (loans) program in Peru. Data come from different sources: two experimental games (trust and public goods game), two individual surveys, one privately conducted, the other publicly, and financial savings and loan data. He also finds that trust questions appear to predict trustworthiness, but not trusting behavior.4

Opposite results are obtained by Bellemare and Kröger (2007) and Fehr et al. (2003). Both studies incorporate a trust game into a survey. Bellemare and Kröger include this game in the CentralPanel survey among 541 Dutch households, of whom 499 participated in the experiment. The survey also contains two trust questions. Fehr et al. (2003) integrate a trust game experiment in their survey among 442 adult individuals living in households within Germany. They include six questions about trust, most of them similar to questions used in Glaeser and Souter (2000). In both studies none of the survey measures of trust are good predictors of trustworthiness in the experiment but, unlike in Glaeser and Souter (2000) and Karlan (2005), some of them are reasonably good in predicting trusting behavior.

Some other studies investigate whether personality characteristics correlate with trustworthiness and trust. For instance, Gunnthorsdottir et al. (2002) use the Machiavellianism scale (Mach-IV), which measures a person’s tendency towards cynicism, manipulativeness and the extent of his/her concerns with conventional

4Most interestingly, in this study observed behavior in the experimental games is correlated with actual behavior in reality. In particular, Karlan finds that trustworthiness as measured by the return ratio in the trust game appears to predict the probability of repayment on loans in the year following the experiment. Moreover, individuals who are trustworthy are also more able to secure loans from their peers.
morality (beliefs that the ends sanctifies the means), to predict behavior in a two-person game similar to the trust game. It appears that the Mach-score does not correlate with trust, but correlates negatively with trustworthiness. Burks et al. (2003) obtain contrasting results. They report an experiment in which University of Minnesota students play both roles in the (otherwise standard) trust game. They relate the outcomes to survey data, which include (among other things) questions on personality characteristics. It appears that the Mach-score does not correlate with trustworthiness, but does correlate negatively with trust.\textsuperscript{5} Ben-Ner et al. (2004) also relate personality factors to observed behavior. In particular, they conduct a two round-dictator game experiment with the roles reversed in the second round. Here, the second round of the game serves to analyze to which extent the second-round senders reciprocate the generosity of first-round senders. They find that reciprocal behavior is only significant when pairs remain the same over the two rounds. It also holds that the more agreeable, neurotic and open the sender is (measured by a personality test), the larger the amount they send in response to the amount they receive. A drawback of Ben-Ner et al. (2004)’s study is that the personality characteristics used, lack a theoretical underpinning why (and in which direction) they determine reciprocal attitudes.

Summing up, the existing studies focus on measuring trust and trustworthiness and do not include questions that directly try to measure reciprocity, i.e. whether someone is willing to reward kind behavior and punish unkind behavior (even when this is costly). Leuven et al. (2005) do include such a question in their survey among Dutch workers, but they do not provide evidence that this question is indeed a satisfactory measure of workers’ propensity to reciprocate (i.e. correlates with actual behavior). Our study attempts to fill this gap.

\textsuperscript{5}The Mach-score is also used to explain choices in other experimental games. Meyer (1992) finds that low Machs are more likely to refuse low offers as second movers in a one-shot ultimatum game than high Machs are. This is interpreted as low Machs being more willing to reciprocate negatively. The experiments of Meyer are hypothetical; choices are not elicited in an incentive compatible way. Carpenter et al. (2005) find the Mach variable to be unimportant in dictator games and marginally significant for the amount offered in ultimatum games.
3 The data and descriptive statistics

3.1 The experiment

In this section we briefly present our experimental design. A more elaborate discussion can be found in Chapter ??.

A total of 144 undergraduate students (42% were females) participated in the experiment. Each student was assigned to one of the 7 sessions. All subjects played a one-shot gift-exchange game either in the role of employer or as worker. Employers decide on the wage, which has to be a multiple of 5 between 0 and 100. Workers simultaneously choose an effort level for each possible wage. More specifically, for all 21 possible wage levels they have to indicate their effort choice, which has to be an integer between 1 and 10. Employers choose the wage without knowing the workers’ effort choices, and workers fill in the wage-effort table without knowing the actual wage set by their employers.

Subjects participated in one of two treatments considered. In the baseline treatment (treatment 1-1) subjects were randomly and anonymously divided into groups of two. In each group, one subject was assigned the role of employer, the other subject the role of worker. They played the gift exchange game as described above. The employer’s monetary payoffs were given by $10e - w + 90$ and the worker’s monetary payoffs equalled $w - c(e) + 90$. Here $e$ denotes the effort chosen by the worker (for the actual wage offered), $w$ is the employer’s wage offer, and $c(e)$ the costs of effort. Note that in the baseline treatment subjects receive the same endowment of 90 points. In the other treatment (treatment 1-4), subjects were randomly and anonymously divided into groups of four. Each group was again divided into two subgroups, and one subgroup was assigned the role of employer while the other subgroup the role of worker. Employers and workers then played the gift exchange game as described above. The employer’s monetary payoffs were given by $5e - w + 45$ and the worker’s monetary payoffs equalled $w - c(e) + 45$. Here $e$ denotes the effort chosen by the worker (for the actual wage offered), $w$ is the employer’s wage offer, and $c(e)$ the costs of effort.

---

6The experiment took place in the CREED laboratory at University of Amsterdam. It was computerized using the z-Tree software developed by Fischbacher (1999).
7This means that we applied the strategy method. A discussion of its advantages and disadvantages is given in Section ?? of Chapter ??.
8Four sessions considered treatment 1-1 (84 participants), the other three sessions concerned treatment 1-4 (60 participants).
9See Table ?? in Chapter ?? for the specific cost schedule used.
groups of five. In each group, one subject was assigned the role of employer, the other four the role of worker. Subjects had to make the same decisions as in treatment 1-1. In particular, the employer had to choose only one wage, i.e. all four workers received the same wage. The workers had to fill-in the wage-effort table without knowing the actual wage and the effort choices made by their co-workers. Monetary payoffs equalled $10(e_1 + e_2 + e_3 + e_4) - 4w + 360$ for the employer, and $w - c(e) + 90$ for each worker. In this treatment, the employer receives four times a worker’s initial endowment.

In both treatments, after all subjects had made their decisions, workers were informed about the actual wage offered by the employer. Employers learned the effort choices of their workers for the wage chosen. Subjects were also informed about their own and their group members payoffs but they were never informed about the choices of subjects in other groups. Finally, each participant had to fill in the ex post questionnaire. Thereafter they were privately paid their own earnings.10

3.2 The survey

Subjects filled out an anonymous short questionnaire. The attitudinal and self-reported behavioral questions were based on the questions used in the papers referred to in Section 2. Specifically, our exit survey included four types of questions/statements:11

1. Reciprocal attitude:

   (a) Own attitude ($Reciprocity1$) - “If someone does something that is beneficial to me, then I am prepared to return a favor, even when this was not agreed upon in advance”;  

10The average earnings were 17.09 euros (including a show-up fee of 5 euros) for about one hour on average.
11Apart from the questions reported, we also included questions about background characteristics like parents’ education, age, field of study and year of enrolment. The complete questionnaire is reproduced in Appendix 4.A.
(b) Expectations about behavior of others (*Reciprocity*2)—“If I do something that is beneficial for someone else, then I expect that person to return a favor”.

2. Trusting attitude

   (a) *Trust1* - “In general, one can trust people”;
   (b) *Trust2* - “In these days you can’t rely on anybody else”;
   (c) *Fair* - “Most people try to take advantage of you if they got a chance”;
   (d) *Help* - “Most of the time people try to be helpful”;
   (e) *Strangers* - “When dealing with strangers it is better to be careful before you trust them”.

3. Past behavior

   (a) *Personal possessions* - “How often does it happen that you lend personal possessions (like CDs, books, bicycle, etc.) to your friends?”;
   (b) *Money* - “How often does it happen that you lend money to your friends?”.

4. Hypothetical choice

   (a) *Tip* - “When you have dinner in a restaurant, how much do you tip the waiter (as percentage of the bill) if a) the service is normal (*Tip*), b) the service is excellent (*Tip*), c) the service is poor (*Tip*)”.¹²

Responses to the first two types of statements are measured on a 5-points scale, ranging from agree strongly (1) to disagree strongly (5). For the questions about past behavior the four answer categories are: almost never (1), sometimes (2), regularly (3) and often (4). In the tipping question the reported tip when the service is normal

¹²We did not stress the one-shot nature of the restaurant visit. However, it seems that tipping behavior is not guided by potential future interactions (see ?), as suggested by Kahnemann et al. (1986).
serves as benchmark. We construct two variables that are meant to capture tipping attitudes. The first one is the difference between tipping an excellent and a normal restaurant service ($\Delta\text{tip} - \text{en}$) and the second one is the difference between tipping a poor and a normal service ($\Delta\text{tip} - \text{np}$).

Two remarks concerning our questions are in order. First, we decided to include self-reported measures of trusting attitudes. Some existing studies (see Glaeser and Souter, 2000; Karlan, 2005) have pointed out that trustworthiness (reciprocity) may be best measured by asking whether the respondent trusts other people. Given that these studies do not include self-reported measures of reciprocity, we do not know whether the significance of trusting questions results omitted variable bias. So, it seems important to also consider trusting measures in our analysis. Second, we decided to include a question about tipping choices. The question is formulated such that it captures the essence of positive and negative reciprocity. Economic analyses on tipping can be found in Bodvarsson and Gibson (2003), Bodvarsson et al. (2003) and Conlin et al. (2003).

### 3.3 Descriptive statistics

In this section we give an overview of our data.\footnote{Here we do not present disaggregated data per treatment. For this see Chapter ??}. We focus first on observed behavior in the experiment. The mean average effort over all possible wages is 3.57. Moreover, mean effort increases with the wage level (see Figure 1). As a measure of observed reciprocal behavior we use the estimated wage coefficient for each individual, i.e. $\partial e/\partial w$. Out of 90 workers, 26 are selfish ($\partial e/\partial w = 0$), 57 behave reciprocal, i.e. choose a monotonic and increasing wage-effort schedule, and 7 subjects display non-monotonic behavior. The average wage offer is 37.72.
Next we focus on self-reported measures. Tables 1-3 present the percentage of respondents that chose the respective categories for questions concerning reciprocal and trusting attitudes as well as past lending behavior. The vast majority of subjects (90.3%) are prepared to return a favor. The distribution of answers to this question is similar to the one reported in Leuven et al. (2005). Also, 63% of respondents expects a favor in return from someone for whom they do something beneficial. So, not all respondents who are reciprocal themselves expect reciprocal behavior from others. This possibly means that some of these subjects do not trust others, or that they are altruistic and do not even require a favor in return. It appears that more respondents have a reciprocal attitude compared to a trusting attitude. Only 36.8% trust others and 78.5% think that it is better to be careful with strangers. Concerning lending behavior, respondents lend more often (or regularly) personal possessions than money. Finally, the size of the tip appears to be related to service quality. On average respondents are willing to give the waiter 9.9%, 5.5%, and 1.2% of the bill if the service is excellent, normal and poor, respectively.
Table 1: Frequencies of the answers to *Reciprocal attitude* questions

<table>
<thead>
<tr>
<th>Reciprocity 1</th>
<th>Reciprocity 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>agree strongly</td>
<td>33 (22.9)</td>
</tr>
<tr>
<td>agree</td>
<td>97 (67.4)</td>
</tr>
<tr>
<td>neither disagree nor agree</td>
<td>7 (4.9)</td>
</tr>
<tr>
<td>disagree</td>
<td>6 (4.2)</td>
</tr>
<tr>
<td>disagree strongly</td>
<td>1 (0.7)</td>
</tr>
</tbody>
</table>

*Note: Percentages are in brackets*

Table 2: Frequencies of the answers to *Trusting attitude* questions

<table>
<thead>
<tr>
<th>Trust 1</th>
<th>Trust 2</th>
<th>Stranger</th>
<th>Fair</th>
<th>Help</th>
</tr>
</thead>
<tbody>
<tr>
<td>agree strongly</td>
<td>2 (1.4)</td>
<td>3 (2.1)</td>
<td>32 (22.2)</td>
<td>12 (8.3)</td>
</tr>
<tr>
<td>agree</td>
<td>51 (35.4)</td>
<td>27 (18.8)</td>
<td>81 (56.3)</td>
<td>60 (41.7)</td>
</tr>
<tr>
<td>neither disagree nor agree</td>
<td>43 (29.9)</td>
<td>41 (28.5)</td>
<td>22 (15.3)</td>
<td>34 (23.6)</td>
</tr>
<tr>
<td>disagree</td>
<td>37 (25.7)</td>
<td>61 (42.4)</td>
<td>8 (5.6)</td>
<td>35 (24.3)</td>
</tr>
<tr>
<td>disagree strongly</td>
<td>11 (7.6)</td>
<td>12 (8.3)</td>
<td>1 (0.7)</td>
<td>3 (2.1)</td>
</tr>
</tbody>
</table>

*Note: Percentages are in brackets*
Table 3: Frequencies of the answers to past behaviour questions

<table>
<thead>
<tr>
<th>Possessions</th>
<th>Money</th>
</tr>
</thead>
<tbody>
<tr>
<td>often</td>
<td>34 (23.6)</td>
</tr>
<tr>
<td>regularly</td>
<td>58 (40.3)</td>
</tr>
<tr>
<td>sometimes</td>
<td>47 (32.6)</td>
</tr>
<tr>
<td>almost never</td>
<td>5 (3.5)</td>
</tr>
</tbody>
</table>

*a Note: Percentages are in brackets*

4 Results

To explore the predictive power of attitudinal questions for subjects’ behavior in the gift exchange game we estimate the following two equations:

$$\left( \frac{\partial e}{\partial w} \right)_i = \alpha_w + \beta_w X_i + \gamma_w T_i + \varepsilon_{iw} \quad i = 1, \ldots, N$$  \hspace{1cm} (1)

$$w_j = \alpha_e + \beta_e X_j + \gamma_e T_j + \varepsilon_{je} \quad j = 1, \ldots, M,$$  \hspace{1cm} (2)

where $\left( \frac{\partial e}{\partial w} \right)_i$ is the estimated wage-effort coefficient for worker $i$ and $w_j$ is the wage offered by employer $j$. We assume that $\left( \frac{\partial e}{\partial w} \right)_i$ and $w_j$ are affected by a vector of observable attitudinal variables $X$, a treatment dummy $T$, and disturbance terms $\varepsilon_{iw}$ and $\varepsilon_{je}$ which are independently and normally distributed. $N$ is the number of subjects with the role of worker and $M$ is the number of subjects with the role of employer.

In total we have 11 attitudinal variables. However, it is unlikely that each variable separately captures a unique attitudinal trait. For example, Trust1 and Trust2 aim at measuring trusting attitudes using subjects opinions on two but closely
related statements. So, in order to avoid problems of multicollinearity in estimating equations (1) and (2) we first use a data reduction method to combine similar attitudes into broader categories (‘factors’). In particular, we perform a factorial analysis using principal components extraction. Formally, in this first step we estimate the following model:

\[ X_g = \mu + \lambda F_g + \nu_g \quad \nu \sim N(0, \psi) \quad g = 1, \ldots, N + M, \]  

where \( X_g \) is a \( p \times 1 \) vector of attitudinal characteristics, \( \mu \) is a vector of means, \( F_g \) is a \( k \times 1 \) vector of factors (\( K < p \)) with \( E[F_g] = 0 \) and \( Cov[F_g] = I \), implying that all factors are orthogonal. \( \lambda \) is a \( p \times k \) matrix of coefficients (factor loadings), \( \nu_g \) is a \( p \times 1 \) vector of disturbance terms that are uncorrelated with the factors. We assume that the matrix \( \psi \) is diagonal and known, implying that all of the correlation among attitudinal characteristics is explained by the common factors. This model is estimated on a sample of \( N+M \) observations.\(^{14}\)

Using principal components extraction with variance maximizing rotation, four factors with an eigenvalue greater than 1 can be extracted (cf. the Kaiser criterion, Kaiser, 1960).\(^{15}\) Once the factor loadings are estimated, we compute the factor scores (i.e., the predicted values of the unobserved factors). Table 4 presents the factor loadings.

The loadings in Table 4 display the correlation between the questions and the factors. The first factor is consistently defined by the five questions that relate to trusting attitudes. We refer to this factor as Trust. By construction, factors are

\(^{14}\)For robustness purposes the analysis is also performed on the two subsamples of employers and workers separately.

\(^{15}\)The same four factors are obtained when using other orthogonal rotation strategies such as quartimax and equamax, or when the analysis is applied to both the subsample of workers and employers. This indicates that the factor structure is stable. Two tests were performed, the Barlett’s test of sphericity and the Kaiser-Meyer-Olkin test of sampling adequacy. The results indicate that the variables are correlated and the principal components analysis applied to the actual set of variables is good.
orthogonal and each represent a different attitudinal trait. Factor 2 mainly consists of the two questions that ask about past lending behavior. We call this factor \textit{Lend}. The third factor is characterized by high loadings on the two reciprocity questions. Hence, we label this factor \textit{Reciprocity}. Finally, the fourth factor is mainly defined by the two variables that capture tipping attitudes. We label factor 4 \textit{Tip}. To interpret the factors easily, the answers to the questions were “positively-recoded”. Therefore, the higher the \textit{Trust} factor, the higher the subject’s trusting attitude. Similarly, the larger the \textit{Reciprocity} factor, the higher the subject’s reciprocal attitude is.
<table>
<thead>
<tr>
<th>Question</th>
<th>Trust</th>
<th>Lend</th>
<th>Reciprocity Tip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reciprocity1</td>
<td>0.115</td>
<td></td>
<td>0.749</td>
</tr>
<tr>
<td>Reciprocity2</td>
<td>0.195</td>
<td>0.661</td>
<td>-0.160</td>
</tr>
<tr>
<td>trust1</td>
<td>0.815</td>
<td></td>
<td>-0.131</td>
</tr>
<tr>
<td>trust2</td>
<td>0.718</td>
<td>0.170</td>
<td>-0.201 0.151</td>
</tr>
<tr>
<td>Stranger</td>
<td>0.541</td>
<td>0.330</td>
<td>-0.120</td>
</tr>
<tr>
<td>Fair</td>
<td></td>
<td>0.732</td>
<td></td>
</tr>
<tr>
<td>Help</td>
<td>0.692</td>
<td>-0.190</td>
<td>0.139</td>
</tr>
<tr>
<td>Possessions</td>
<td></td>
<td>0.822</td>
<td></td>
</tr>
<tr>
<td>Money</td>
<td>0.153</td>
<td></td>
<td>0.826</td>
</tr>
<tr>
<td>ΔTip-en</td>
<td></td>
<td>-0.171</td>
<td>0.864</td>
</tr>
<tr>
<td>ΔTip-pn</td>
<td>-0.148</td>
<td>-0.161</td>
<td>0.475 -0.521</td>
</tr>
</tbody>
</table>

*Note:* Principal components extraction with varimax rotation and Kaiser normalization. Loadings with an absolute value less than 0.1 are replaced by blanks; those highest in each equation are bold-faced.

It is reassuring to find that the four factors correspond exactly with the a priori classification of the survey questions. In fact, the same group of questions was previously answered by 248 students who had subscribed online for participating in experiments. In particular, subjects who enrolled themselves in our database for the first time (up to 6 months before the experiment), were asked to answer a short questionnaire, which would yield them 5 euros. (From this group of students only 68 participated in the current experiment.) This short survey included, among other questions, all the questions used here. We also applied the factor analysis to the “ex-ante” survey data and exactly the same factors were obtained and ranked in the same order. Using the factor loadings that resulted from the factor analysis for the ex-ante
questionnaire we constructed factor scores for all participants in the present experiment (144 subjects). Comparing these factors with those obtained from applying the factor analysis to the exit survey we observed a high and significant correlation between the first three factors. Only the factors mainly composed of the tip questions are not significantly correlated. The factor structure thus appears robust.

Turning to our second step, we estimate a linear regression equation (see equation 1) with the individual worker’s wage-effort coefficient $\frac{\partial c}{\partial w}$ as dependent variable. For the workers that are classified as being selfish this coefficient equals zero. Although our main interest lies in workers’ behavior, we also estimate regression equation 2 with the employer’s wage offer as dependent variable. In both equations the constructed factor scores based on the estimated factor loadings presented in Table 4 are used as regressors, together with a dummy indicating the 1-4 treatment.

16Having 68 subjects who answered the questionnaire twice, we compared their factor scores for the four factors in both samples using a signrank test. Results suggest no significant differences between ex-ante and ex-post factors. This is in line with Bellemare and Kröger (2007), who claim that attitudes are not affected by behavior.
Table 5: Worker’s inferred type $\frac{\partial e}{\partial w}$ and employer’s wage equations

<table>
<thead>
<tr>
<th></th>
<th>$\frac{\partial e}{\partial w}^{(1)}$</th>
<th>wage$^{(2)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.0571</td>
<td>33.244*</td>
</tr>
<tr>
<td></td>
<td>(0.0067)</td>
<td>(4.336)</td>
</tr>
<tr>
<td>Trust</td>
<td>0.0000</td>
<td>-11.084*</td>
</tr>
<tr>
<td></td>
<td>(0.0042)</td>
<td>(4.04)</td>
</tr>
<tr>
<td>Lend</td>
<td>-0.0030</td>
<td>-0.029</td>
</tr>
<tr>
<td></td>
<td>(0.0043)</td>
<td>(4.304)</td>
</tr>
<tr>
<td>Reciprocity</td>
<td>0.0082**</td>
<td>-0.373</td>
</tr>
<tr>
<td></td>
<td>(0.0043)</td>
<td>(3.851)</td>
</tr>
<tr>
<td>Tip</td>
<td>-0.0060</td>
<td>1.667</td>
</tr>
<tr>
<td></td>
<td>(0.0041)</td>
<td>(3.72)</td>
</tr>
<tr>
<td>Treatment 1-4</td>
<td>-0.0010</td>
<td>2.309</td>
</tr>
<tr>
<td></td>
<td>(0.0096)</td>
<td>(8.471)</td>
</tr>
</tbody>
</table>

N 90 workers 54 employers

Notes: (1) Feasible weighted least squares estimates, with weights equal to $1/\sqrt{\hat{\sigma}^2 + \omega_i^2}$ where $\hat{\sigma}^2$ is the estimated homoscedastic variance of the error term in equation 1, and $\omega_i^2$ is the variance of the error term in the first stage equation; (2) OLS estimates with robust standard errors. * (**) indicates significance at 1% (5%) level.

Table 5 contains our main results. The second column reveals that questions about trust, past lending behavior and tipping behavior do not correlate with worker’s actual reciprocal behavior. The only factor that reaches significance is composed of questions that ask subjects directly about their reciprocal attitude.\(^{17}\) The higher the

\(^{17}\)Excluding the workers who display erratic behavior, i.e. whose effort choices do not follow a monotonic pattern, we get a significant coefficient (at 5% level) for the Tip factor. Restricting the analysis to the 68 subjects that participated in both the ex ante survey and in the experiment, similar
Reciprocity factor, the larger a worker’s slope coefficient of the wage-effort schedule. Looking at the two reciprocity questions in isolation, it appears that only the first one correlates significantly with the worker’s type. It does not matter whether a worker agrees or disagrees with Reciprocity2. The slope coefficient is essentially predicted by Reciprocity1, as one would a priori expect.

We also used the same four factors to explain the wage offers made by the employers in the experiment. As the third column reveals, only questions that ask subjects about their trusting behavior appear statistically significant. Again, this is in line with what one a priori would expect. An employer is more likely to offer a high wage if s/he is more trusting that the worker will reciprocate in kind.

5 Conclusion

In this chapter we investigate whether attitudinal and behavioral survey questions explain observed behavior in a gift exchange laboratory experiment. To that purpose we estimate two equations. The first one concerns workers’ behavior, i.e. the dependent variable is the estimated wage-effort coefficient for subjects in the role of workers. The second one concerns employers’ behavior. Here the dependent variable is the wage offer.

Questions that ask subjects directly about their concerns for reciprocity are the best predictors of observed reciprocity. The questions about trusting attitude do not correlate with reciprocity. However, these questions appear to be a good predictor of employers’ wage offers. Because in the gift exchange game the size of the wage offer reflects (among other things) the trusting attitude of employers, it appears that our trust questions do correlate with trust. Our results thus contrast with those of Glaeser and Souter (2000) and Karlan (2005).

Results are found, but no significant coefficients are found for the regressions reported in Table 5 due to small numbers (40 workers and 28 employers, respectively) of observations.
We think that our results give some hints as to which type of questions should be used in future research that aims at measuring the impact of reciprocity. Certainly, more research needs to be done. For instance, an issue that seems worth investigating is whether reciprocal experimental behavior and survey attitudinal questions do predict actually observed reciprocal behavior in “the real world”.

References


Appendices

A. Summary of the instructions

B. Ex-post survey

Questions about the experiment

1. If I could choose my own role in another session of this experiment, I would rather be:
   - An employer □
   - A worker □

2. Could you please indicate to what extent you agree or disagree with the following statements.
   
   (a) I think it is fair that a worker who gets a high wage should put in more effort.
       - agree strongly □
       - agree □
       - neither agree nor disagree □
       - disagree □
       - disagree strongly □

   (b) (Only for workers) When deciding on my effort choice for the different possible wages, I realized what the consequences would be for the earnings of the employer.
       - agree strongly □
       - agree □
       - neither agree nor disagree □
       - disagree □
       - disagree strongly □

   (c) (Only for employers) When I was deciding on the wage, I expected that a worker who gets a higher wage would exert more effort.
       - agree strongly □
       - agree □
       - neither agree nor disagree □
       - disagree □
       - disagree strongly □

Other questions

At the end of this experiment we would like to ask you to fill in this questionnaire. There are no correct answers, we are only interested in your opinion about a number of issues.
3. Could you please indicate to what extent you agree or disagree with the following statements.

(a) If someone does something that is beneficial to me, then I am prepared to return a favor, even when this was not agreed upon in advance.
   - disagree strongly
   - disagree
   - neither agree nor disagree
   - agree
   - agree strongly

(b) If I do something that is beneficial for someone else, then I expect that person to return a favor.
   - disagree strongly
   - disagree
   - neither agree nor disagree
   - agree
   - agree strongly

(c) In general, one can trust people.
   - disagree strongly
   - disagree
   - neither agree nor disagree
   - agree
   - agree strongly

(d) In these days you can’t rely on anybody else.
   - disagree strongly
   - disagree
   - neither agree nor disagree
   - agree
   - agree strongly

(e) When dealing with strangers it is better to be careful before you trust them.
   - disagree strongly
   - disagree
   - neither agree nor disagree
   - agree
   - agree strongly

(f) Most people try to take advantage of you if they got a chance.
   - disagree strongly
   - disagree
   - neither agree nor disagree
   - agree
   - agree strongly

(g) Most of the time people try to be helpful.
disagree strongly □
disagree □
neither agree nor disagree □
agree □
agree strongly □

4. How often does it happen that:

(a) You lend personal possessions (like CDs, books, bicycle etc.) to your friends
    often □
    regularly □
    sometimes □
    almost never □

(b) You lend money to your friends
    often □
    regularly □
    sometimes □
    almost never □

5. When you have dinner in a restaurant, how much do you tip the waiter (as percentage of the bill) if:

    service is normal (i.e. like it should be) □ %
    service is excellent □ %
    service is poor □ %

6. What is the highest level of education completed:

(a) of your father?
    primary school □
    lower vocational (vmbo) □
    secondary school (havo/vwo) □
    medium vocational (mbo) □
    higher vocational (hbo) □
    university (hbo) □

(b) of your mother?
    primary school □
    lower vocational (vmbo) □
    secondary school (havo/vwo) □
    medium vocational (mbo) □
    higher vocational (hbo) □
    university (hbo) □

The first time you signed up for an experiment through the CREED-website you supplied some personal data and potentially also filled in the web-questionnaire. We would like to link your choices to the data in our database. Please note, we are not interested in linking your choices to your name, but rather to your background characteristics and, if available, to your answers to the web-questionnaire.
In case you do not have an objection against this, please fill in your student ID number below (when you are not a student at UvA, please use the customary abbreviation of the institution you are studying, followed by your student ID number).

Student ID number:

Some questions about your background

Finally we would like to ask you some questions about your background. In case you filled in your student ID number above, you can skip the questions below.

Please fill in your age:  

male  
female  

Field of study/ Department:  
UvA: Department of Economics and Econometrics  
UvA: Department of Humanities  
UvA: Law Department  
UvA: Behavioral Sciences: Psychology  
UvA: Behavioral Sciences: Non-Psychology  
UvA: Medicine / Dental Surgery  
UvA: Physics, Mathematics and Computer Science  
UvA: Different university  
UvA: Higher vocational education (HBO)  
UvA: Otherwise  

Year of enrollment:  

This is the end of the questionnaire. We kindly ask you to remain seated and keep quiet until all subjects have finished the questionnaire.

Thank you very much for participating in our experiment!