

Decomposing Trust and Trustworthiness

Nava Ashraf, Iris Bohnet and Nikita Piankov[†]

What motivates people to trust and be trustworthy? Is trust solely “calculative,” based on the expectation of trustworthiness, and trustworthiness only reciprocity? Employing a within-subject design, we ran investment and dictator game experiments in Russia, South Africa and the United States. Additionally, we measured risk preferences and expectations of trustworthiness. Expectations of trustworthiness account for most of the variance in trust, but unconditional kindness also matters. Variance in trustworthiness is mainly accounted for by unconditional kindness, while reciprocity plays a comparatively small role. There exists some heterogeneity in motivation but people behave surprisingly similarly in the three countries studied.

Key Words: Trust, reciprocity, social preferences, cross-cultural experiments

JEL Classification: C72, C91

[†]Ashraf: Harvard Business School; Bohnet: Kennedy School of Government, Harvard University; Piankov: Analysis Group, Boston. Corresponding author: Iris Bohnet, 79 JFK Street, Cambridge, MA02139, phone: (617) 495-5605, fax: (617) 496-5747, email: Iris_Bohnet@Harvard.edu.

1. Introduction

Recently, one of the authors found herself standing in line in a supermarket, with a basket full of groceries but no wallet. She decided to ask the person behind her whether he could lend her \$20. He did and gave her his business card so that she could return the money. No questions asked.

In such a situation, one party, the trustor, makes himself vulnerable by taking a course of action that creates incentives for the other party, the trustee, to exploit him. Such behavior is commonly referred to as trust. It is assumed that the trustor would not choose the risky course of action if he did not expect the trustee to honor trust. Indeed, a large body of work in many disciplines assumes that trust is “calculative,” based on the *expectation of trustworthiness* (e.g., Rotter 1980, Williamson 1993, Hardin 2002). Trustworthiness, in turn, is typically assumed to be *reciprocity* in behavioral economics (for surveys, see Camerer 2003, Ostrom and Walker 2003).¹

While the dominant view, these assumptions have recently been challenged by various scholars in economics, political science and psychology (e.g., Kramer 1999, Mansbridge 1999, Dufwenberg and Gneezy 2000, Cox 2004).² Rather than being motivated by expectations of return, the lender in the supermarket may have given the money because he enjoys trusting others. And the borrower may have returned the money, not because she wanted to reciprocate

¹ Reciprocity is defined as rewarding kindness with kindness and punishing unkindness with unkindness. Behavior conditional on one’s counterpart’s behavior or intentions has been formalized by Rabin (1993), Charness and Rabin (2002), Cox, Friedman, and Gjerstad (2005), Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (forthcoming). It has been examined experimentally in a large number of studies (for a survey, see Fehr and Schmidt 2002).

² Kramer (1999, p. 573) argues that “trust needs to be conceptualized not only as a calculative orientation toward risk, but also a social orientation toward other people and toward society as a whole.” Mansbridge (1999: 294) notes the importance of “altruistic trust”: “To qualify as altruistic trust my move must... be motivated by a conscious or unconscious intent to benefit the other or an intent to uphold a principle that in the long run usually benefits others.” Recent neuroscientific evidence from Kosfeld et al. (2005) also suggests that trust is not only a belief: Trust is related to oxytocin levels, a hormone typically associated with love and pro-social behavior.

the lender's kindness but rather because she derived personal satisfaction from being trustworthy. Such *unconditional kindness* could be due to social preferences such as altruism (Andreoni and Miller 2002), inequity aversion (Fehr and Schmidt 1999, Bolton and Ockenfels 2000) or quasi-maximin preferences (Charness and Rabin 2002) or due to the psychological benefits, a "warm glow" (Andreoni 1990), which an individual derives from being kind to others. Social preferences and internalized norms could induce people to trust and be trustworthy—despite the fact that trust often does not "pay" monetarily, and trustworthiness never "pays" monetarily in one-shot interactions. In his survey on experimental outcomes, Camerer (2003, p. 87) writes: "The fact that the return to trust is around zero seems fairly robust."

In fact, the data suggest that if people perceived trust only as an investment decision under uncertainty—as suggested by the name of the game most widely used to measure trust, the "investment game" (Berg, Dickhaut and McCabe 1995)—they would keep making bad investments. Already in the first study by Berg et al. (1995), trustors lost money on average, a finding that by now has been replicated in many experimental studies (for surveys, see Camerer 2003, Cardenas and Carpenter 2005 and Greig and Bohnet 2005). Even if subjects were informed of previous trustors' and trustees' behavior, the trust level did not decrease. This "social history" treatment was first used by Berg et al. (1995) and then replicated by Ortmann, Fitzgerald and Boeing (2000) who found that even if trustors were graphically shown that trust typically "does not pay," their trust was not affected. Ortmann et al. also elicited trustors' expectations of return. Only 29 percent of their trustors in the social history treatment expected to make money in this game.

To better understand the role social preferences play in trust situations, behavior in the investment game has been compared with giving in otherwise identically structured dictator games (Kahneman, Knetsch and Thaler 1986) in recent studies. Dufwenberg and Gneezy (2000) and Cox (2004) find that only relatively small fractions of the money sent by the trustor and returned by the trustee are due to expectations of trustworthiness or reciprocity. Both studies use a between-subjects design. They assume that expectation-based trust and reciprocity are the additive difference between what is sent in the dictator and the investment games.

Our paper contributes to this discussion but does not make any such strong assumptions. Rather, we use a within-subject design where the same person participates in a dictator and an investment game (in varied order). We test whether the two decisions are related, controlling for expectations of trustworthiness and attitudes to risk.³ Used often in experimental psychology but little in experimental economics, within-subject designs automatically control for individual differences. Given the substantial heterogeneity in social preferences observed (e.g., Andreoni and Miller 2002), we feel more comfortable with a within- rather than a between-subjects design.

As a robustness check, we run our study in three different countries, Russia, South Africa and the United States. To connect our research with earlier studies on trust,⁴ we also confronted subjects with the standard attitudinal trust question used in the World Values Survey and the General Social Survey: “Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?” We focus on single-play

³ We will refer to attitudes to risk to describe our subjects' behavior but are aware that we are dealing with relatively small stakes here and thus may want to describe behavior by attitudes to losses instead (Rabin 2000).

⁴ See the literature started by Glaeser, Laibson, Scheinkman and Soutter (2000).

anonymous interactions. While excluding important contextual features that may affect trust and trustworthiness in naturally occurring settings, they measure “pure trust and trustworthiness” and provide a benchmark for comparisons with more complex environments.

We aim to better understand what game people are playing when confronted with the investment game. Our design allows us to measure the relative importance of beliefs (reciprocity) as compared to unconditional kindness and whether there is heterogeneity in the motivation accounting for observed trust (trustworthiness) behavior. Our paper is organized as follows. In the next Section II, we present the experimental design. Section III presents our conceptual framework. Section IV discusses the experimental results and Section V concludes.

2. Experimental Design

The experiment consisted of an introduction page and five parts, which was common knowledge. After having signed a consent form, experimental participants received the introduction page detailing some general experimental rules. Subjects were informed that the experiment will consist of five parts, that they will receive the instructions for each part separately, that they will remain anonymous during the experiment (i.e., identified by code numbers) and that they will be randomly paid in the end. After having read the introduction, subjects were randomly assigned to two different rooms, one for the trustors and one for the trustees.

In Part I, all subjects were asked to complete a questionnaire, collecting information on demographic characteristics and trust attitudes.⁵ In Part II, all subjects played a dictator game

⁵ The experimental instructions are available from the authors upon request. As Glaeser et al. (2000), we elicited trust attitudes before we measured trust behavior experimentally. This allows us to examine whether the—surprising—lack of a correlation between trust attitudes and trust behavior in the Glaeser et al. study was due to subjects knowing each other. Their investment game measured trust in a specific, identified individual, which may not be correlated with general trust attitudes.

in the role of the dictator. The dictator was asked to allocate a fixed endowment S between herself and a recipient who did not receive any money. The dictator earned $S-X$ and the recipient earned X . In Part III, all subjects played a triple dictator game in the role of the dictator. The only difference to the standard dictator game is that any amount X given to the recipient was tripled by the experimenter. Accordingly, the dictator's earnings were as before, $S-X$, but the recipient's earnings were $3X$.⁶

In Part IV, subjects played the trust game—in room A in the role of the trustor and in room B in the role of the trustee. The only difference between the trust game and the triple dictator game was that the trustee could send back any amount of money Y between 0 and $3X$. The trustor's earnings thus were $S-X+Y$ while the trustee earned $3X-Y$. We used the strategy method where the trustees had to decide on a contingent action for every possible amount sent by the trustors. In addition to indicating how much they wanted to send, we also asked trustors to report what they expected to get back.⁷

In Part V, all subjects had to indicate for six risky choice tasks whether they preferred the gamble or the certain amount. They could choose to bet on a 50%-chance of winning 300CU or nothing or to accept a certain amount that varied between 40CU and 140CU in the six choice tasks. The more people prefer the sure thing to the gamble, the more risk averse they are.

⁶ While everyone played the game in the role of the dictator, subjects were informed that at the end of the whole experiment one of them would be randomly chosen to be the dictator and one the recipient.

⁷ To decrease complexity in an already quite complex experiment, we chose not to reward subjects for accuracy of expectations. The evidence about whether payment increases accuracy or not is not conclusive. In a recent extensive study on the relationship between beliefs and actions, Costa-Gomes and Weizsäcker (2003) conclude that the effects of the belief elicitation procedure on actions is mostly insignificant. There is general agreement that subjects are more likely to take their counterpart's incentives into account when beliefs are elicited than when making action choices only (e.g., Croson 2000, Costas-Gomez and Weizsäcker 2003).

The sequence in which subjects participated in Parts II, III, IV and V was varied. In half of our sessions, subjects played these parts in order; in the other half, subjects were confronted with Part IV first before participating in Parts II, III and V. Each part of the experiment was conducted like a “standard experiment”: The instructions for each part, consisting of one decision form, were distributed one at a time. After subjects had read the instructions, the experimenter read the instructions aloud. Participants were then invited to ask questions in private (hardly any questions were asked). Subjects then had to write down their decision, fold the decision form and put it into a box, which was passed around. Completed decision forms remained in the box in front of the room until the end of the experiment. The same procedure was repeated for each part.

The endowment S was 100CU (currency units) in our experiments. Adjusted for purchasing power parity, this meant $S = \$100$ in the United States, $S = 1000$ Roubles in Russia and $S = 400$ Rands in South Africa.⁸ In the trust game, the strategy method was implemented as follows: trustors had to choose between 11 possible amounts to be passed on to the trustee (0, 10, ..., 100 CU) and trustees had to indicate for each possible positive amount how much they would return. We did not restrict subjects’ choices in the dictator games. They could send any amount between 0 and 100 CU.

Subjects were paid randomly at the end of the whole experiment; they did not learn about any results during the experiment. More specifically, for each game in parts II to IV, two people (a dictator and a recipient or a trustor and a trustee) were randomly selected and matched at the end of the experiment to be paid according to their choices in the

⁸ We chose denominations such that the monetary incentives relative to subject income and living standards were approximately equal across countries. The experiments were conducted in 2001. The average lunch in the student cafeteria cost \$5 in Boston, 50 Roubles in Moscow and 20 Rands in South Africa.

corresponding game. For the individual decision task in part V, one person was randomly paid according to his or her choice.

The experiments were conducted with 359 college students: 118 students from universities in Moscow, Russia, 129 students from universities in Capetown, South Africa, and 112 students from universities in Boston, United States. We ran four experimental sessions in each country, two with the dictator games first and two with the trust game first. The experimenters who ran experiments in Russia and in South Africa also ran one session in the United States. No experimenter effects could be found.⁹ The experiment took about one hour and thirty minutes. A show-up fee of 10CU was paid and subjects earned on average an additional 22CU.

A non-standard experimental format was required to test our hypotheses—a within-subject design to control for subjects' social preferences and expectations; a three-country design to allow for heterogeneity in behavior and motivation more globally than just among American subjects; the strategy method to learn about each trustee's return function; and high stakes with random payment to make conducting such a long and complex experiment feasible. We address the first two aspects of our design by controlling for order and country effects econometrically. There is only little evidence on how the latter two design aspects might affect behavior. For example, Brandts and Charness (2000) suggest that the strategy method does not affect behavior differently from the choice method and Laury (2002) presents evidence supporting the validity of the random-choice payments method. She found that subjects take (high) stakes at their stated value and do not scale-down to account for random

⁹ One of the experimenters fluent in both English and Russian translated the instructions into Russian. They were back-translated by an external translator not familiar with this research. In order to ensure equivalence of experimental procedures across countries, we followed Roth et al. (1991) on designs for multinational experiments and controlled for experimenter, currency and language effects to the best of our ability.

payment. While we control for design effects to the best of our ability, we cannot exclude the possibility that our design affects behavior differently than other, more standard, designs.

3. Conceptual Framework

Trust is defined as the amount sent, X , and trustworthiness as the amount returned, Y , divided by the amount received, $3X$, that is $Y/3X$. $E(Y/3X)$ indicates trustors' expectations of the fraction returned. TDG_{give} is the amount sent in the triple dictator game and DG_{give} the amount sent in the dictator game. Table A.1 in the Appendix provides the specific definitions of all our variables. We estimate the following equations:

$$X = \alpha + \beta * E(Y/3X) + \gamma * TDG_{give} + \delta * \text{controls} \quad (1)$$

$$Y/3X = \alpha + \beta * 3X + \gamma * DG_{give} + \delta * \text{controls} \quad (2)$$

We compare the following hypotheses for trust:

H1_T: Trust is only based on expectations of trustworthiness.

We take as evidence for $H1_T$ if X is related to the expected fraction returned, $E(Y/3X)$, but not to the amount sent in the triple dictator game, TDG_{give} . Further support for $H1$ is provided if the constant is close to zero, indicating that trustors do not derive any intrinsic benefits, i.e. a warm glow, from trusting.

H2_T: Trust is only based on unconditional kindness.

$H2_T$ is supported if X is related to the amount sent in the triple dictator game, TDG_{give} , but not to the expected fraction returned $E(Y/3X)$. Additional support for this hypothesis is provided if the constant is significantly positive, suggesting that people derive a warm glow from trusting.

For trustworthiness, we compare:

H1_{TW}: Trustworthiness is only based on reciprocity.

We take as evidence for H1_{TW} if $Y/3X$ is related to the amount received, $3X$, but not to the amount sent in the dictator game, DGgive.¹⁰

H2_{TW}: Trustworthiness is only based on unconditional kindness.

H2_{TW} is supported if $Y/3X$ is strongly related to the amount sent in the dictator game, DGgive, but not to the amount received, $3X$.

To be able to compare our results with earlier work (Dufwenberg and Gneezy 2000, Cox 2004), we use behavior in the dictator and the triple dictator games to assess our subjects' unconditional kindness. We acknowledge that this is a relatively crude step as we assume that kindness is unconditional on what one's counterpart's actions are (or are expected to be) as well as on the payoff-possibilities following those (expected) actions.

The approach is of particular concern for trustworthiness as the relative income positions of the trustee and the trustor may differ dramatically from the relative income positions of the dictator and the recipient. In contrast, in the trust and the triple dictator game situations, both, the trustor and the dictator, start with CU100 and the trustee and the recipient with zero (plus a show-up fee of 10CU in both cases). To test for the robustness of our results, we thus run two sets of regressions for trustworthiness, one where we include dictator game giving directly as a proxy for unconditional kindness and one where we impose a particular social preference profile on all trustees, namely fixed distributional preferences.¹¹ The latter provides us with a measure of how much subjects would remit in the trust game based solely on the distributional preferences they exhibited in the dictator game.

¹⁰ A positive intercept can be interpreted similarly as above but we do not discuss it further as trustees could not return a positive amount when they received zero.

¹¹ Fixed distributional preferences are compatible with inequity aversion, see Fehr and Schmidt (1999) and Bolton and Ockenfels (2000).

Holding a trustee's distributional preference constant means that her payoff must be related to her trustor's payoff in the same way her payoff as a dictator is related to her recipient's payoff. We call this the *predicted (distributional preference-based) remit function*:

$$\text{Ratio in Trust Game} = \frac{3X - Y}{100 - X + Y} = \frac{100 - DGgive}{DGgive} = \text{Ratio in Dictator Game.}$$

Accordingly, a trustee would have to return

$$Y = \frac{2X * DGgive}{100} + DGgive + X - 100.$$

The ratio to be remitted, predicted based on fixed distributional preferences, is $Y/3X$:

$$\frac{Y}{3X} = \left(\frac{2X * DGgive}{300X} \right) + \frac{DGgive + X - 100}{3X}.$$

Since the predicted ratio above can be negative, but $Y < 0$ is not a feasible choice for the trustee, we assign zero to all predicted ratios that are negative:

$$\text{predictedremit} = \max \left\{ \left(\frac{2X * DGgive}{300X} \right) + \frac{DGgive + X - 100}{3X}, 0 \right\}.$$

Experimental evidence suggests substantial heterogeneity in subjects' preference profiles. Andreoni and Miller (2002), for example, found evidence for the existence of at least three different types of players in dictator games when examining how dictators respond to the price of giving. Accordingly, we are somewhat uneasy about imposing one specific preference profile on all our subjects. We examine the degree of heterogeneity in our subjects' response modes by comparing our subjects' behavior in the dictator and the triple dictator games. Table 1 summarizes our subjects' preference profiles.

Table 1: Distribution of choices in the dictator games

TDGgive=DGgive		TDGgive<DGgive		TDGgive>DGgive
Give=0CU	Give>0CU	DGgive=50CU TDGgive=25CU	DGgive≠50CU or TDGgive≠25CU	
14%	16%	15%	24%	31%

14 percent of our subjects are selfish and do not send anything in either game. 16 percent send the same positive amount in the DG and the TDG. 39 percent send less in the TDG than in the DG, trying to (approximately) preserve whatever distributional preference they exhibited in the DG (for example, by sending 50CU in the DG and 25CU in the TDG). 31 percent respond to the decrease in the price of giving in the TDG by sending more in the TDG than in the DG. A majority of our subjects, 53 percent, behave according to a fixed distributional preference profile. While we clearly notice substantial heterogeneity, a fixed distributional preference profile describes our subjects better than any other profile.

4. Experimental Results

We first present an overview of our behavioral data and trustors' expectations of return and then proceed with the specific results related to our hypotheses.

Overview

On average, our trustors send 45CU of their endowment of 100CU to their trustees (N=179), and our trustees return 27 percent of the money received to their trustors (N=1790). As we used the strategy method, we have ten times as many data points for the trustees than for the trustors because the latter indicated how much they would return for each possible positive amount sent to them. (We cluster the data by individuals in the regressions.) The method also keeps us from losing data due to those trustees who would not have received

anything. In the dictator game, subjects send 25CU and in the triple dictator game 24CU (N=358) on average.

We first note that the mean amounts sent and returned in our games are surprisingly close to the standard results in trust and dictator games despite the fact that our design substantially differs from earlier designs. Camerer (2003) reports in his survey of experimental results that typically dictators send about 20 percent while trustors send about 50 percent of their endowment and trustees return about the amount that trustors sent to them, i.e. about one third of the tripled amount on average. The triple dictator game has not been studied widely. Cox (2004) finds that subjects send 36 percent of their endowment on average, which is somewhat more than what we find.

Figures 1a-1d in the Appendix present the distribution of choices in all games for the three countries. A Mann-Whitney and a Kolmogorov-Smirnov test reveal no significant differences in the means and the distributions of choices of Americans, Russians and South Africans or of men and women in the three games.¹² The biggest difference in behavior can be found between white and nonwhite South Africans in the trust game, with nonwhite trustors sending 33CU and white trustors sending 52CU to their trustees (Mann-Whitney U-test, $p < 0.01$, Kolmogorov-Smirnov test, $p < 0.05$).

159 of our 175 trustors¹³ send a positive amount to their trustees. Out of those who trust, 38% expect to get back less than they sent, 26% expect to break even and 36% expect back more than they sent. Expectations are surprisingly well calibrated and close to the

¹² The experimental evidence on gender is not conclusive. However, in most dictator games, women tend to send more than men and in most trust games, women tend to be less trusting but more trustworthy than men (for a survey, see Croson and Gneezy 2004). For more information on the break-down of our data, please see our working paper (Ashraf, Bohnet and Piankov 2004).

¹³ Out of the 179 trustors in our full sample, 2 did not answer the question to measure expectations of return, and 2 expected back a sum that they couldn't possibly get back because it was above 3 times the amount they sent.

average amount returned, the social norm, in a given country. Comparing average amounts expected back for each possible amount sent with the actual returns received, we find that the only significant outlier is full trust, i.e., $X=100\text{CU}$. Trustors expect to get back about 150CU but trustees only return about 100CU on average (M-W U-test, $p<0.01$).

Result 1: Trust: Expectations of trustworthiness or unconditional kindness?

In order to better understand what is driving the variation in our sample, we run multivariate regressions where in addition to expectations of return and unconditional kindness, we control for attitudes to risk, a number of demographic variables, attitudes to trust and order effects.

Table 2 below presents the regressions for the trustors. A basic model of demographic characteristics, trust attitudes and risk preferences explains little of the variation observed (15% of the total variation, see Column 1). This basic model suggests that risk preferences do not matter for trust,¹⁴ that South African non-whites trust less than others and that there is an order effect: those who play the dictator game first, send significantly less in the trust game ($X=37\text{CU}$) than those who play the trust game first ($X=52\text{CU}$). Column 2 presents a regression where we add the expected proportion returned, $E(Y/3X)$, to the set of explanatory variables. The coefficient on expectations is highly significant statistically and economically: an increase in expected proportion returned by 1% increases the amount sent by 1.1CU, on average. The inclusion of expected fractions returned substantially reduces the order effect, both in magnitude and in significance. When people play the dictator game first, they expect to get

¹⁴ Studies focusing on the relationship between people's willingness to take risk and trust behavior report mixed results. Eckel and Wilson (2004) found no relationship between risk-taking and trust behavior for American subjects while Schechter (forthcoming) reported a strong correlation between attitudes to risk and behavior in a trust game run in Paraguay. Bohnet and Zeckhauser (2004) suggest that in addition to attitudes to risk, attitudes to betrayal have to be taken into account in the trust game.

back a slightly though not significantly smaller fraction ($E(Y/3X)=0.28$) than when they play the trust game first ($E(Y/3X)=0.32$).¹⁵ Column 3 includes the amount sent in the triple dictator game in addition to the controls. TDGgive is highly significant. Those who send 1CU more in the triple dictator game tend to send 0.7CU more in the trust game.¹⁶

In Column 4, we find that trust decisions are related to both expectations of return and unconditional kindness, although the coefficients on both, proportion expected back and especially TDGgive, decrease in magnitude when the two are included together. In all specifications, the constant is significantly positive, suggesting that trustors derive satisfaction from trusting independent of amounts expected back. Column 4 also suggests a persistent race effect in South Africa: nonwhites trust less than whites in South Africa but not in the U.S. Our findings on race are in line with earlier results on race: It is the groups that historically felt discriminated against, which are less likely to trust (Alesina and LaFerrara 2002). At the beginning of the 21st century, such discrimination is more prevalent in South Africa than in the United States.

¹⁵ These beliefs appear to be well-calibrated: trustees give back slightly less in the trust game when they played the dictator game first ($Y/3X=0.26$) than when they played the trust game first ($Y/3X=0.28$).

¹⁶Note that 24 of our trustors send less in the trust game than in the triple dictator game, suggesting that expectation-based trust is not just the additive difference between what is sent in the trust and the triple dictator game. Comparing men and women, expectations matter more for women's and unconditional kindness more for men's trust (see Ashraf et al. 2004 for a more detailed analysis).

Table 2. Determinants of amount sent in the trust game¹⁷

	(1)	(2)	(3)	(4)
Prop. expected back		112.774 (8.969)***		101.291 (9.005)***
TDGgive			0.708 (0.119)***	0.379 (0.093)***
Risk aversion	-1.702 (1.580)	-1.054 (1.115)	-2.165 (1.432)	-1.368 (1.066)
Dictator 1st (yes=1)	-17.227 (4.987)***	-6.399 (3.618)*	-15.410 (4.524)***	-6.530 (3.449)*
South Africa	12.728 (9.128)	9.196 (6.438)	6.345 (8.330)	6.143 (6.183)
Russia	5.910 (7.882)	0.751 (5.570)	7.773 (7.140)	2.272 (5.322)
US non-white	-6.346 (8.979)	-0.963 (6.342)	-0.609 (8.184)	1.556 (6.078)
SA non-white	-28.447 (8.997)***	-18.110 (6.394)***	-19.241 (8.289)**	-14.240 (6.169)**
Gender (male=1)	-5.128 (5.405)	-3.112 (3.813)	-6.079 (4.895)	-3.826 (3.639)
Age group (1-6)	-0.480 (4.167)	-1.046 (2.937)	-1.137 (3.773)	-1.340 (2.800)
Econ. situation (1-6)	-3.985 (2.805)	-3.627 (1.977)*	-3.149 (2.542)	-3.217 (1.887)*
Economics major	1.520 (6.628)	-4.354 (4.694)	4.178 (6.015)	-2.335 (4.502)
# of organizations	-0.674 (1.951)	-1.393 (1.376)	-0.083 (1.769)	-1.003 (1.316)
Trust strangers	1.781 (6.787)	-0.582 (4.786)	-1.013 (6.160)	-1.835 (4.573)
Constant	74.334 (16.562)***	39.410 (11.997)***	55.618 (15.317)***	32.960 (11.546)***
Observations	168	168	168	168
R-squared	0.15	0.58	0.31	0.62
Adjusted R-squared	0.09	0.55	0.25	0.59

The dependent variable is X, the amount sent in the trust game. Standard errors in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%.

¹⁷ Our sample size for the trust regressions is 168. From our complete sample of 179, we lose 7 subjects because of missing observations for one of the independent variables, 2 subjects who expected to get back more than 3 times what they sent, and we exclude 2 outliers. These are a Russian man and a South African black woman who both send 10CU and expect to get back 30CU. If we include them, expectations of return become less correlated with amount sent while unconditional kindness gains in importance. While this would strengthen the message of this paper, we prefer our results not to be significantly influenced by these two outliers.

Result 2: Trustworthiness: Reciprocity or unconditional kindness?

A trustee returns money to the trustor if she is motivated by reciprocity and/or unconditional kindness. The positive relationship between amounts received and amounts remitted has mostly been interpreted as an indicator for reciprocity, negative and positive, in the literature so far (e.g., Camerer and Fehr 2004). However, without further analysis the conclusion that a positive slope of the return function is an indicator for reciprocity seems only warranted for unconditionally selfish players who send nothing in the dictator game. Any positive fraction that they return in the trust game cannot be due to innate kindness. For everyone else, however, a positive relationship between amounts received and fractions remitted warrants further analysis.

As trustors send more money to the trustees (and keep less for themselves), trustees have to return proportionally more money to live up to their distributional preferences. An egalitarian type with fixed distributional preferences, for example, would not remit any money for small amounts sent but would remit 50% if she received 300CU. Figure 1 illustrates how much an egalitarian type who sends 50CU in the dictator game would have to remit in the trust game according to the predicted remit function introduced above, i.e., if she were motivated by her distributional preferences only. The theoretical preference-based remit function is compared with the experimentally observed remit function.

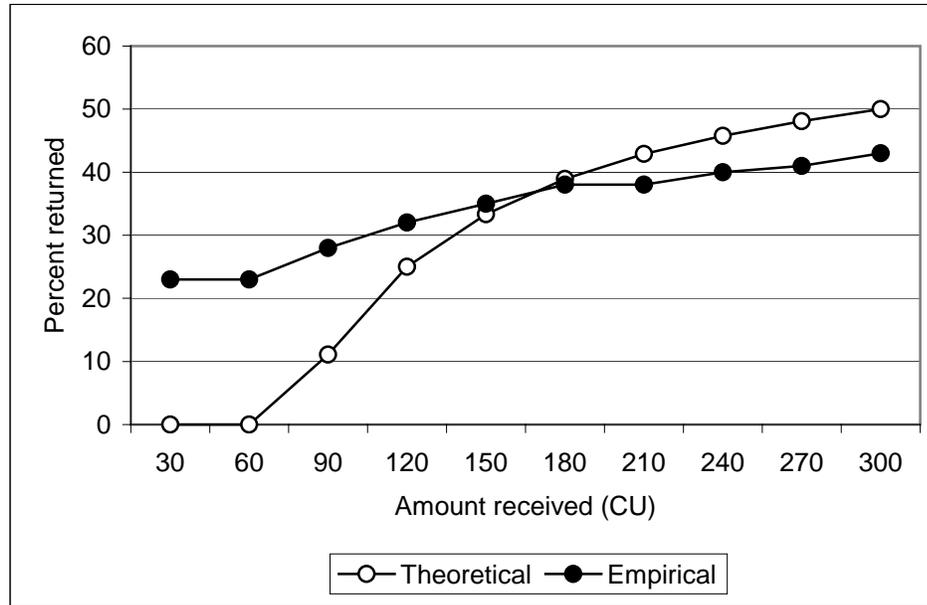


Figure 1: Fractions remitted by egalitarian types-based on their predicted (theoretical) and observed (empirical) remit functions.

Figure 1 shows that egalitarian types remit more than predicted by the preference-based theoretical remit function when sent small amounts and less when sent large amounts. This pattern does not suggest that these players are motivated by reciprocity.

In order to differentiate between the two motives more precisely, we run a regression. We take it as an indicator of reciprocity if the percent returned, $Y/3X$, increases with the amount received, $3X$, controlling for kindness. To test for whether the specification of kindness affects results, in Table 3, we include the amounts sent in the dictator game directly into our regressions in Column 3 and use a fixed distributional preference profile in Column 4.

Table 3 presents the regression results for the trustees. Our control variables in Column 1 explain only 7% of the variation in our sample. Two control variables are significant. Russians remit more than others and “Trust strangers” is related to the percentage returned, suggesting that generally, the trust question picks up trustworthiness rather than trust. This finding is line with Glaeser et al.'s (2000) results for Harvard undergraduates. When our main

explanatory variables – amount received in the trust game, amount given in the dictator game and predicted distributional preference – are included, all have high statistical and economic significance. In Column 2, we find that one extra CU received increases the percentage of the money returned by 0.05%, which means that when a subject receives the full 300CU, she returns about 16% of the amount received. Column 3 shows that our subjects return 0.3% more of the amount received for every CU they send in the dictator game, meaning that an “egalitarian” person who splits the money equally in the dictator game will return 15% more than an egoist who keeps all the money in the dictator game.

When reciprocity (Amount received) and kindness (DGgive) are combined in Column 5, both variables maintain their levels and significance, suggesting that these two factors coexist in the decision process. However, if DGgive is replaced by the predicted distributional preferences, the effects of reciprocity disappear (Column 6). This suggests that the observed positive slope of percentage returned need not be based on people’s willingness to reward trust by being kind in return. They may only care about their distributional preferences.

Independent of the specification, including the behavioral variables increases the variation explained to a bit more than 20%. Overall, we reject both hypotheses. However, in contrast to trust, unconditional kindness accounts for most of the variance explained while reciprocity plays a comparatively small role.

Table 3. Determinants of percentage returned in the trust game¹⁸

	(1)	(2)	(3)	(4)	(5)	(6)
Received (3X)		0.00052 (0.00006)***			0.00052 (0.00006)***	0.00010 (0.00009)
DGgive			0.003 (0.001)***		0.003 (0.001)***	
Predicted distr. pref.				0.442 (0.051)***		0.416 (0.066)***
Risk aversion	-0.009 (0.009)	-0.009 (0.009)	-0.010 (0.008)	-0.009 (0.008)	-0.010 (0.008)	-0.009 (0.008)
Dictator 1st (yes=1)	-0.023 (0.024)	-0.023 (0.024)	-0.014 (0.021)	-0.016 (0.021)	-0.014 (0.021)	-0.016 (0.021)
South Africa	0.028 (0.038)	0.028 (0.038)	0.024 (0.036)	0.026 (0.036)	0.024 (0.036)	0.026 (0.036)
Russia	0.073 (0.033)**	0.073 (0.033)**	0.061 (0.030)**	0.060 (0.030)**	0.061 (0.030)**	0.061 (0.030)**
US non-white	0.011 (0.035)	0.011 (0.035)	0.033 (0.036)	0.022 (0.035)	0.033 (0.036)	0.021 (0.035)
SA non-white	-0.019 (0.038)	-0.019 (0.038)	-0.023 (0.035)	-0.025 (0.035)	-0.023 (0.035)	-0.025 (0.035)
Gender (1=male)	-0.004 (0.024)	-0.004 (0.024)	0.003 (0.022)	0.002 (0.022)	0.003 (0.022)	0.002 (0.022)
Age Group (1-6)	0.001 (0.021)	0.001 (0.021)	0.001 (0.019)	0.001 (0.019)	0.001 (0.019)	0.001 (0.019)
Economic situation (1-6)	-0.001 (0.011)	-0.001 (0.011)	0.003 (0.011)	0.004 (0.011)	0.003 (0.011)	0.004 (0.011)
Economics major	0.044 (0.031)	0.044 (0.031)	0.026 (0.028)	0.027 (0.029)	0.026 (0.028)	0.028 (0.029)
# of organizations	-0.002 (0.008)	-0.002 (0.008)	-0.002 (0.007)	-0.001 (0.007)	-0.002 (0.007)	-0.001 (0.007)
Trust strangers	0.085 (0.028)***	0.085 (0.028)***	0.054 (0.025)**	0.057 (0.025)**	0.054 (0.025)**	0.059 (0.025)**
Constant	0.257 (0.073)***	0.172 (0.074)**	0.174 (0.070)**	0.196 (0.068)***	0.089 (0.071)	0.184 (0.069)***
Observations	1770	1770	1770	1770	1770	1770
R-squared	0.07	0.13	0.18	0.22	0.24	0.23
Adjusted R-squared	0.06	0.12	0.18	0.22	0.24	0.22

The dependent variable is Y/3X, the amount returned divided by the amount received. Standard errors in parentheses; observations are clustered by individual. * significant at 10%; ** significant at 5%; *** significant at 1%

¹⁸ Our sample size for the trustworthiness regressions is 177. From our complete sample of 179, we lose 2 subjects because of missing observations for one of the independent variables.

5. Conclusions

In his book on “Behavioral Game Theory”, Colin Camerer (2003) suggests that one of the “Top Ten Open Research Questions” is the following: “What game do people think they are playing?” Our results suggest that many people may play a different game than researchers thought they were playing when confronted with the “investment game.” Only 36 percent of the 159 trustors who decided to send any money in our “investment game” expected to make money.

Trust is based on beliefs of trustworthiness and on unconditional kindness; trustworthiness is related to unconditional kindness and reciprocity. Expectations of trustworthiness account for most of the observed variance in trust and unconditional kindness for most of the explained variation of trustworthiness behavior.

We would not have been able to estimate the relative importance of expectations of return and reciprocity as compared to unconditional kindness, had we only compared average behavior in a between-subject design. Our experiments are a first attempt at better understanding what motivates people to do the things they do, by including preferences, as revealed by behavior, as explanatory variables. We believe that many other economic experiments could benefit from a combination of within- and between-subjects designs.

Our design allowed us to solve one of the important trust puzzles, namely that people trust even though hardly anyone makes money by doing so. We found that generally, people are aware of this. They trust even though they know it does not pay monetarily. They enjoy being kind to others, even to anonymous strangers. If these psychological returns of trust are taken into consideration, people may not make a bad investment when trusting.

References

- Alesina, A. and La Ferrara, E. (2002) "Who Trusts Others?" *Journal of Public Economics*. 85, 207-34.
- Andreoni, J. (1990). "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving?" *Economic Journal*. 100, 464-77.
- Andreoni, J., and Miller, J. (2002). "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica*. 70, 737-53.
- Ashraf, N., Bohnet, I., and Piankov, N. (2004). "Decomposing Trust and Trustworthiness." Working paper, Kennedy School of Government, Harvard University.
- Berg, J., Dickhaut, J., and McCabe, K.A. (1995), "Trust, Reciprocity, and Social History." *Games and Economic Behavior*. 10, 290-307.
- Bohnet, I. and Zeckhauser, R. (2004). "Trust, Risk and Betrayal." *Journal of Economic Behavior and Organization*. 55, 467-485.
- Bolton, G. and Ockenfels A. (2000). "A theory of equity, reciprocity and competition," *American Economic Review*. 90, 166-93.
- Brandts, J. and Charness, G. (2000). "Hot vs. Cold: Sequential Responses and Preference Stability in Experimental Games," *Experimental Economics*. 2, 227-38.
- Camerer, C.F. (2003). *Behavioral Game Theory*. Princeton: Princeton University Press.
- Camerer, C.F. and Fehr, E. (2004). "Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists." In: Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr and Herbert Gintis (eds.), *Foundations of Human Sociality*. Oxford: Oxford University Press.
- Cardenas, J.C. and Carpenter J. (2005). "Three Themes on Field Experiments and Economic Development." Working paper: Middlebury College.
- Charness, G. and Rabin, M. (2002). "Understanding Social Preferences With Simple Tests." *Quarterly Journal of Economics*. 117, 817-69.
- Costa-Gomez, M.A. and Weizsäcker, G. (2003). "Stated Beliefs and Play in Normal-Form Games." Mimeo.
- Cox, J.C. (2004) "How To Identify Trust and Reciprocity." *Games and Economic Behavior*. 46, 260-281.
- Cox, J.C., Friedman, D., and Gjerstad, S. (2005). "A Tractable Model of Reciprocity and Fairness." Working paper: University of Arizona.
- Croson, R. (2000). "Thinking Like a Game Theorist: Factors Affecting the Frequency of Equilibrium Play." *Journal of Economic Behavior and Organization*. 41, 299-314.
- Croson, R. and Gneezy, U. (2004). "Gender Differences in Preferences." Working paper: Wharton School, University of Pennsylvania.
- Dufwenberg, M. and Kirchsteiger, G. (2004). "A theory of sequential reciprocity." *Games and Economic Behavior*. 47, 268-298.
- Dufwenberg, M. and Gneezy, U. (2000). "Measuring Beliefs in an Experimental Lost Wallet Game." *Games and Economic Behavior*. 30, 163-82.
- Eckel, C.C. and Wilson, R.K. (2004). "Is Trust a Risky Decision?" *Journal of Economic Behavior and Organization*. 55, 447-466.
- Falk, A. and Fischbacher, U. "A Theory of Reciprocity." *Games and Economic Behavior*. In Press.
- Fehr, E. and Schmidt, K. (1999). "A Theory of Fairness, Competition and Cooperation." *Quarterly Journal of Economics*, 114, 817-68.

- Fehr, E. and Schmidt, K. (2002). "Theories of Fairness and Reciprocity -- Evidence and Economic Applications." In: Mathias Dewatripont, Lars Peter Hansen, and Steven J. Turnovsky (eds.), *Advances in Economics and Econometrics - 8th World Congress, Econometric Society Monographs*. Cambridge: Cambridge University Press.
- Glaeser, E.L., Laibson, D.I., Scheinkman, J.A. and Soutter, C.L. (2000) "Measuring Trust." *Quarterly Journal of Economics*. CXV, 811-46.
- Greig, F. and Bohnet, I. (2005). "Is There Reciprocity in a Reciprocal-Exchange Economy? Evidence from a Slum in Nairobi, Kenya." Working paper: Kennedy School of Government, Harvard University.
- Hardin, R. (2002). *Trust and Trustworthiness*. New York: Russell Sage.
- Kahneman, D., Knetsch, J., and Thaler, R. (1986). "Fairness as Constraint on Profit Seeking: Entitlements in the Market." *American Economic Review*. 76, 728-41.
- Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U., and Fehr, E. (2 June 2005). "Oxytocin Increases Trust in Humans," *Nature*. 435, 673-76.
- Kramer, R. (1999). "Trust and Distrust in Organizations: Emerging Perspectives, Enduring Questions." *Annual Review of Psychology*. 50, 569-98.
- Laury, S. (2002). "Pay One or Pay All: Random Selection of One Choice for Payment." Mimeo, Georgia State University.
- Mansbridge, J. (1999). "Altruistic Trust," In: Mark E Warren (ed.), *Democracy and Trust*. Cambridge: Cambridge University Press, pp. 290-309.
- Ostrom, E., and Walker, J. (2003). (eds.), *Trust and Reciprocity*. New York: Russell Sage.
- Ortmann, A., Fitzgerald, J. and Boeing, C. (2000). "Trust, Reciprocity, and Social History: A Re-examination." *Experimental Economics*. 3, 81-100.
- Rabin, M. (1993). "Incorporating Fairness into Game Theory and Economics." *American Economic Review*. 83, 1281-1302.
- Rabin, M. (2000). "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica*. 68, 1281-92.
- Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M., and Zamir S. (1991). "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study." *American Economic Review*. 81, 1068-95.
- Rotter, J. (1980). "Interpersonal trust, trustworthiness and gullibility." *American Psychologist*. 35, 1-7.
- Schechter, L. (forthcoming). "Traditional Trust Measurement and the Risk Confound: An Experiment in Rural Paraguay." Working paper: University of California at Berkeley, *Journal of Economic Behavior and Organization*.
- Williamson, O.E. (1993). "Calculativeness, trust and economic organization." *Journal of Law and Ethics*. 36, 453-86.

Appendix

Table A.1: Variable definitions

Concept	Variable Name	Variable Description	Range of values
Trust behavior	Send Trust	Amount trustor sends in Trust Game	0 to 100CU
Trustworthiness behavior	Percent remitted	Amount trustee returns, as a proportion of amount received	0 to 1
Warm-glow kindness for Trust	TDGgive	Amount subject sends in Triple Dictator Game	0 to 100CU
Warm-glow kindness for Trustworthiness	DGgive	Amount subject sends in Dictator Game	0 to 100CU
Social preferences for Trustworthiness	Predicted distr. pref.	Distributional preference subject exhibits in Dictator Game	0 to 1
Expectations of return	Prop. expected back	Amount trustor expects back, as a proportion of amount sent by trustor	0 to 1
Reciprocity	Received	Amount trustee receives in Trust Game, equals "Send Trust" * 3	0 to 300CU
Attitude to risk	Risk aversion	Number of risky gambles rejected in favor of sure thing by subject	0 to 6
Order effect	Dictator 1 st	Whether the subject played the dictator game prior to the trust game	0 = No, 1 = Yes
South Africa	South Africa	Whether the subject participated in the experiment in South Africa	0 = No, 1 = Yes
Russia	Russia	Whether the subject participated in the experiment in Russia	0 = No, 1 = Yes
Race	Race	Whether the subject is non-white	0 = White, 1 = Non-white
Race in US	US non-white	Whether the subject is a non-white person in the US	0 = No, 1 = Yes
Race in SA	SA non-white	Whether the subject is a non-white person in South Africa	0 = No, 1 = Yes
Gender	Gender	Gender of the subject	0 = female, 1 = male
Age	Age group	Age of the subject	1 to 6 (1 = "Under 20", 6 = "Over 60")
Economic situation	Economic situation	Economic well-being of the subject	1 to 6 (1 = "Poor", 6 = "Wealthy")
Education	Economics major	Whether the subject majors in economics	0 = No, 1 = Yes
Organizational membership	# of organizations	Number of organizations the subject belongs to	Integer value
Trust attitude towards strangers	Trust strangers	"Generally speaking, which of the following people do you feel you could trust not to cheat you?"	1 – if subject trusts one of the following: members of one's religion, citizens of one's country, or foreigners. 0 – if subject does not trust any of the above groups

Appendix

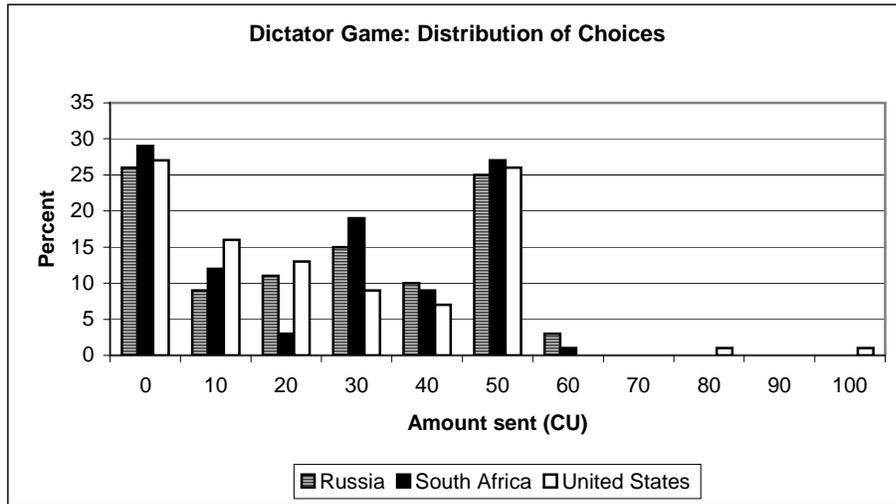


Figure 1a: Dictator game choices.

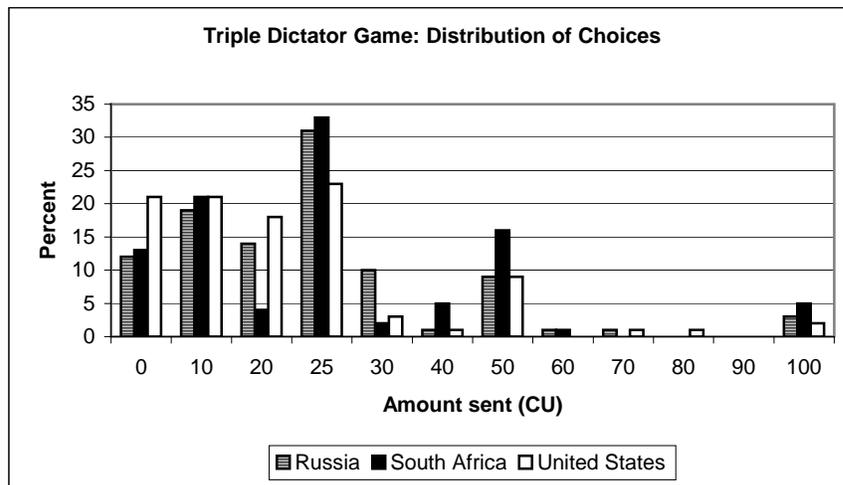


Figure 1b: Triple dictator game choices.

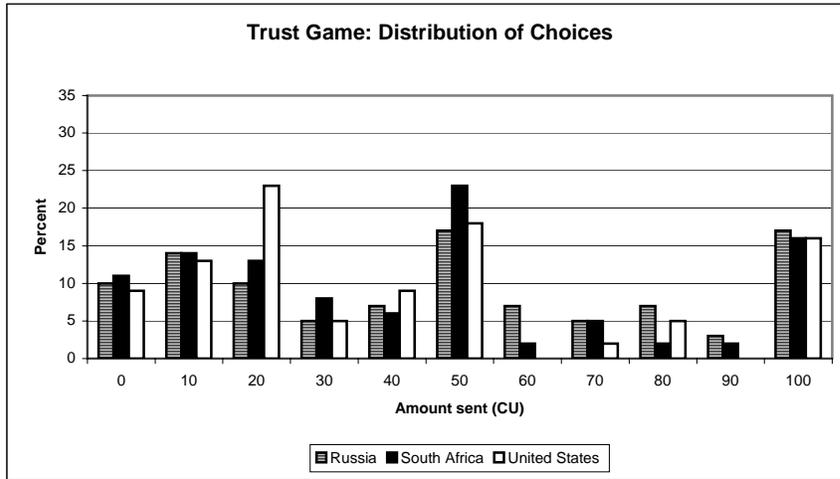


Figure 1c: Trust game choices - trustors.

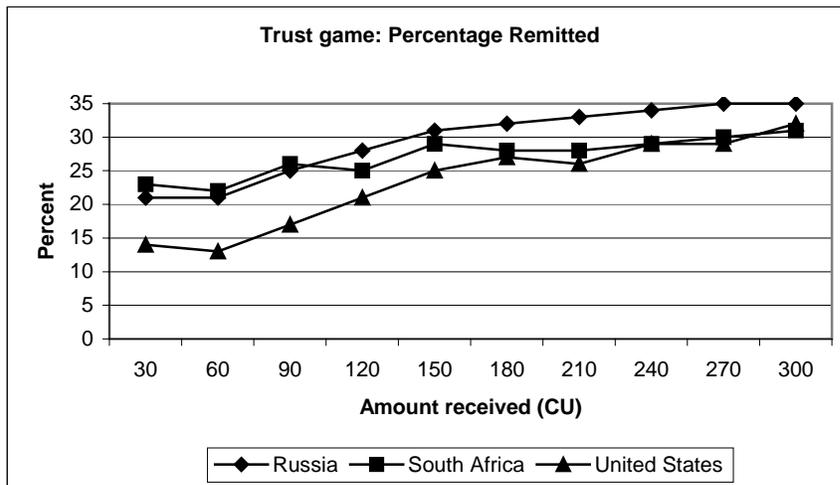


Figure 1d: Trust game choices - trustees.