Public Policy
Discussion Papers
Federal Reserve Bank of Boston

No. 05-4

Deciding to Distrust

Iris Bohnet and Stephan Meier

Abstract:

We employ experiments to illustrate one factor contributing to the lack of distrust in the recent corporate scandals: Trust rather than no trust was the default. People are more trusting when the default is full trust than when it is no trust. We introduce a new game, the distrust game (DTG), where the default is full trust and find that in it, trust levels are higher than in the Berg, Dickhaut, and McCabe (1995) trust game (TG), where the default is no trust. At the same time, trustworthiness levels are lower in the DTG than in the TG. Agents (second movers) punish distrust more in the DTG than the lack of trust in the TG, but principals (first movers) do not correctly anticipate this. The distrust game produces more efficient outcomes than the trust game but also more inequality: Principals end up much worse than their agents in the DTG.

Keywords: trust, reciprocity, reference points, behavioral economics, experimental economics

JEL Classifications: C72, C91

Iris Bohnet is Associate Professor of Public Policy at the John F. Kennedy School of Government at Harvard University. Stephan Meier is Economist at the Federal Reserve Bank of Boston's Research Center for Behavioral Economics and Decision-Making. Their e-mail addresses are: iris_bohnet@harvard.edu and stephan.meier@bos.frb.org, respectively.

This paper, which may be revised, is available on the web site of the Federal Reserve Bank of Boston at http://www.bos.frb.org/economic/ppdp/index.htm.

This version: October 2005

I. Introduction

The recent scandals in big corporations pose many questions. Among them is the puzzle of why shareholders and stakeholders, including legislators, did not cease to trust the companies and their accounting firms earlier. In hindsight, it certainly would have paid to distrust earlier. Many people lost their jobs, their pensions, and their investments. *The Economist* reports that investors lost over $900 billion in thirty big scams between 1997 and 2004 (2005, p. 65). Clearly, informational asymmetries, misaligned incentives, and psychological biases created substantial hurdles. Bazerman, Morgan, and Loewenstein (1997) pointed out many of these concerns before Enron's and WorldCom's implosions, and Bazerman and Watkins (2004) discussed why no action was taken at the time. We propose an additional contributing factor: Trust rather than no trust was the default. People are more trusting when the default is trust than when it is no trust.

This paper examines why people do not *distrust* enough. In contrast to the vast literature on trust, where the default is no trust, we study the effects of a default of full trust. We believe that in many principal-agent relationships characterized by incomplete contracts, principals start out trusting their agents. For example, patients generally believe in their doctor's integrity, parents typically trust their children's teachers, and clients mostly have confidence in their attorneys—and in many cases rightly so. However, compared with a reference point of no trust, principals will not use as much scrutiny and will be overly optimistic about their agent's trustworthiness. The trust default will lead people to trust more than they would have, had the default been no trust, often leaving them worse off than if they had not trusted at all.

We employ experiments to study the behavioral consequences of changing the default in a trust relationship. The investment or trust game (TG) by Berg, Dickhaut, and McCabe (1995) has become the paradigm for measuring trust. In this trust game, as in all other games used to examine trust experimentally, the default is no trust.[1] At the beginning, the principal is endowed

[1] Alternative games used to study trust include binary-choice trust games (Camerer and Weigelt 1988; Kreps 1990), the gift-exchange game (Fehr, Kirchsteiger, and Riedl 1993) and various versions of social dilemma games.

with a certain amount of money; she then decides how much she[2] wants to entrust to her agent. Any amount given to the agent is automatically multiplied by some factor $k > 1$ to capture the efficiency-increasing potential of trust. Agents then decide how much of the amount received to return to their principal. The amount the principal gives to the agent is commonly taken to measure trust, and the fraction returned out of the amount given, trustworthiness.

Traditional economic models based on selfish material preferences and common knowledge thereof assume no trustworthiness and no trust. Various behavioral regularities, such as social preferences (for example, Fehr and Schmidt 1999; Bolton and Ockenfels 2000; Andreoni and Miller 2002), predict outcomes away from the traditional equilibrium. However, as traditional models, such outcome-based social preference models do not predict that the default affects behavior.

A large literature on the relevance of reference points in psychology and behavioral decision theory, building on Kahneman and Tversky (1979), suggests that defaults matter. In our new game, the *distrust game* (DTG), the default is full trust. Agents start with the whole endowment, and principals decide in the first stage how much to take from this endowment. The amount taken is divided by k to capture the idea that distrust has efficiency losses. Then, the agent decides how much out of the remaining money to return to his principal.

Reference-point dependent theories (for example, Kahneman and Tversky 1979; Thaler 1980; Samuelson and Zeckhauser 1988) predict that this change in the default matters. Not trusting in the TG, or the *lack* of trust, is not identical to ceasing to trust in the DTG, or the *loss* of trust. The former corresponds to an act of omission and the latter to an act of commission. For agents, the behavioral decision-theoretic predictions are clear: They should return less in the distrust game than in the trust game for given trust levels, because of loss aversion (Kahneman and Tversky 1979) and reciprocal preferences (Rabin 1993), which punish acts of commission more heavily than acts of omission.

For principals, the story is more complicated. If principals have rational expectations of agent behavior, understanding that the expected returns from trusting (that is, giving in the TG

[2] For ease of understanding, we refer to the principal as "she" and the agent as "he."

and not taking in the DTG) are smaller in the distrust game than in the trust game, then they should trust less in the former than in the latter. However, if principals are not able to walk in their agents' shoes, their expectations may not be accurate and may lead to similar trust levels in the distrust game as in the trust game. Psychological research on perspective-taking suggests that achieving an unbiased perception of other people's preferences and beliefs is hard (see, for example, Samuelson and Bazerman 1985; Neale and Bazerman 1991). For instance, people tend to underestimate the impact of the endowment effect on others' preferences (van Boven, Dunning, and Loewenstein 2000, 2003). In addition, principals' own loss aversion induces them to trust more in the DTG than in the TG.

Our results support the behavioral hypothesis that the default in a trust relationship matters. We find significantly less trustworthiness in the distrust game than in the trust game. Our results suggest that this is due to reciprocity: Agents perceive loss of trust very differently from lack of trust and punish acts of commission more severely than acts of omission. However, principals do not fully anticipate this. Although they expect differences in the slopes of the trustworthiness functions, namely, that agents respond more negatively to distrust in the DTG than to lack of trust in the TG, they do not accurately anticipate that agents in the distrust game are less trustworthy for every level of trust, including full trust. Principals are more optimistic about returns in the DTG than in the TG for high levels of trust (and less optimistic in the DTG than in the TG for low levels of trust), leading them to trust their agents even more in the DTG than in the TG. Notably, 40 percent of principals entrust their agents with all their money in the DTG, while only 16 percent of the principals fully trust in the TG.

Inaccurate perspective-taking in the distrust game affects principals' earnings significantly: They lose money in the distrust game, while they break even in the trust game, on average. Overall, the default of full trust creates substantial inequality between principals and agents. In the distrust game, agents earn about three times as much as principals; in the trust game, agents earn about 1.5 times as much as principals. The magnitude of these effects depends on the parameters used in our experiments. Yet, they remind one of the consequences experienced by many of the workers and investors after the implosion of the large corporations.

The paper proceeds as follows: Section II presents the experimental design and introduces the distrust game. Section III formulates our behavioral hypotheses. The results are presented in section IV. Section V discusses the results and concludes.

II. Experimental Design

The experiments are designed to investigate the effect of changing the default in a trust situation. The control treatment is a standard two-person, anonymous one-shot trust game (Berg, et al. 1995). In the trust game (TG), both principals and agents receive an endowment of \$10. In the first stage, principals can decide whether to keep their endowment or send $x \in [0,1,...,10]$ (in whole dollars) to their agents. x is automatically tripled by the experimenter (k=3). In the second stage, agents can return any amount $y \in [0,1,...10+3x]$ to their principals.

In the distrust game (DTG), principals are initially endowed with \$0, and agents with \$40. In the first stage, the principals can decide whether they want to leave their agents with the endowment and await their decision, or take $z \in [0,3,...,30]$ from them ($z = 3(10-x)$). z is divided by three by the experimenter. For example, if a principal decides to claim \$30, her payoff after stage 1 is $z/3 = \$10$ and her agent's payoff $\$40 - z = \$10$. In the second stage, the agents can "return" any amount $y = [0,1,...,40-z]$ to their principals. Importantly, the two games, the TG and the DTG, differ only in the starting point, no trust or full trust, but not in their payoff space.

Subjects were randomly assigned to be Person X (principal) or Person Y (agent) (see the Appendix for a sample of the instructions). After reading the instructions out loud, questions were answered in private. Then, all subjects completed a quiz to make sure they understood the decision situation. Finally, in the first stage, Persons X (principals) decided how much to give to their agents in the TG or how much to take from their agents in the DTG. The decision sheet included a full payoff table to make the consequences of each decision clear. Persons X's decisions were randomly distributed to Persons Y (agents). In the second stage, Persons Y decided how much of their payoffs after stage 1 they wanted to return to their Persons X. This

concluded the experiment—Each Person X learned about her Person Y's decision at the end of the study.

After subjects had completed the experiment, we handed out a post-experimental questionnaire. In it, we collected information on the possible factors motivating behavior in our two games. Agents' behavior may be motivated by outcome-based preferences, such as other-regarding concerns[3] and loss aversion (Kahneman and Tversky 1979), and/or intention-based preferences, that is, reciprocity (Rabin 1993) and reciprocal responses to commissions and omissions (Kahneman and Tversky 1982). To test the relative importance of outcome-based versus intention-based motives, we asked agents to indicate their trustworthiness in two decision scenarios in the questionnaire. We used the strategy method to collect information on all possible amounts agents could have received in stage 1. In scenario 1, a human principal determined the outcome in stage 1 in the DTG and the TG, much as in the experimental situation. In scenario 2, the stage 1-outcome was determined by a random mechanism, employing a *random distrust game* (RDTG) and a *random trust game* (RTG). The RDTG (RTG) differs from the DTG (TG) only in that the first decision is determined by a random mechanism. Each agent's decisions affect another person's payoffs in all games. Thus, in the random versions of the games, agents in effect become dictators who allocate some (randomly determined) amount of money between themselves and their recipient-principals (for the standard dictator game, see Kahneman, Knetsch, and Thaler, 1986).

Principals' behavior may also be related to their social preferences and loss aversion. In addition, they may base their decisions on their expectations of trustworthiness. Principals were asked to indicate their expectations about agents' trustworthiness in the questionnaire for each possible level of trust, again using the strategy method. We elicited subjects' expectations after their decisions to avoid any influence of this elicitation on subjects' behavior before they learned about their agent's choice (see Croson 2000).

[3] For reviews of theories and empirical evidence on social preferences, see Fehr and Schmidt (2002) and Meier (2004).

To get a sense of the relevance of outcome-based preferences, we asked principals to report how much they would give (not take) in the first stage of a TG (DTG) if there were no second stage. Principals could give either from their endowment of \$10, which was tripled (TG condition), or take up to \$30 from the other party's endowment of \$40, which was divided by three (DTG condition). The agent could only accept this allocation. Thus, subjects played a (triple) giving dictator game in the first case and a (triple) taking dictator game in the second case (Ashraf, Bohnet, and Piankov 2004).

The experiments were conducted with students from various universities in the greater Boston area. One hundred and thirty four subjects participated in our experiments, 64 in the TG and 70 in the DTG. We ran six experimental sessions. Subjects received their final earnings (a show-up fee of \$10 plus their earnings in the experiment) in private at the end of a session. On average, subjects earned \$24 (including the show-up fee) for a 30–40 minute experiment.

III. Behavioral Hypotheses

In the TG, trust is defined as the amount sent, x. Equivalently, in the DTG, trust is defined as the amount not taken, $x = 10 - (z/3)$. Trustworthiness is defined as the amount returned divided by the amount sent/not taken for positive amounts of trust, that is, y/x in the TG and $y/(10-(z/3))$ in the DTG. If subjects care only about their own monetary payoffs and this is common knowledge, the predictions are straightforward for both games: The agent does not return any money. Rational principals anticipate agents' behavior and do not send any money to their agents (in the TG) or they take the maximum amount (in the DTG). In both games, parties would end up with \$10 each.

Evidence from earlier trust experiments does not support this prediction. Subjects show nontrivial levels of trust and trustworthiness. Average trust levels range from 30 percent of the endowment in a slum in Nairobi (for an overview of results in developing countries, see Cardenas and Carpenter 2005; Greig and Bohnet 2005) to about 50 percent of the endowment in many developed countries (for an overview, see Camerer 2003). Trustworthiness varies between 54 percent of the amount given in Kenya (Ensminger 2000) and 128 percent of the amount given

in Zimbabwe (Barr 2003). In developed countries, agents return on average the amount given, that is, 100 percent (Camerer 2003).

Agents' behavior in previous trust games has been found to be driven by both types of social preferences, outcome-based fairness aspects of the resource distribution and intention-based reciprocity. Principals' behavior has been found to be related to their expectations of trustworthiness and their social preferences (for example, McCabe, Rigdon, and Smith 2003; Ashraf et al. 2004; Cox 2004). Bohnet and Zeckhauser (2004) show that in trust relationships intentions are of crucial importance. People care not only about the distribution of material payoffs but also about how the distribution evolved,that is, what the intentions of the other party were.

Both aspects of social preferences might be sensitive to a change in reference points: the frame might influence (1) the perception of the distribution of outcomes and (2) the perception of the process that led to the outcomes. In the following, we discuss how such changes in perception might influence trustworthiness and trust.

Trustworthiness

Agents may not respond to the different reference points in the DTG and the TG. Our null hypothesis is that trustworthiness levels in the DTG and the TG do not differ. In contrast, the behavioral hypothesis suggests:

> **Hypothesis 1:** Trustworthiness levels for given levels of trust are lower in the DTG than in the TG.

We expect a change in the reference points to lead to less trustworthiness for each level of trust in the DTG than in the TG. We offer two conjectures for why trustworthiness might be lower in the DTG than in the TG.

> **Conjecture 1a:** Differences in trustworthiness levels between the DTG and the TG are due to agents' loss aversion.

If people have reference-dependent preferences, a change in the reference point should influence behavior. Various studies support the notion that a reference point frames outcomes as

either losses or gains. People decide differently in the loss domain than in the gain domain. They assign more weight to potential losses than to potential gains (Kahneman and Tversky 1979). Due to loss aversion, people favor the status quo over an alternative (Samuelson and Zeckhauser 1988) and value an item more if they are endowed with it than if they are not (Thaler 1980).

The trust game and the distrust game differ in their reference points. Agents' default is \$10 in the TG and \$40 in the DTG. Independent of principals' actions, compared with the reference point, any amount returned is a loss for the agent in the DTG. In the TG, agents may preserve the status quo if they return $y \leq 3x$. Due to agents' loss aversion, trustworthiness for any given level of trust should be lower in the DTG than in the TG.

> **Conjecture 1b:** Differences in trustworthiness levels between the DTG and the TG are due to reciprocal responses to omission and commission.

Agents' trustworthiness is often attributed to reciprocal preferences. The more benign the agent perceives the principal's intentions of an action to be, the more he will reward trust (Rabin 1993; McCabe, et al. 2003). In our game, this implies that trustworthiness should be increasing in trust.[4]

The change of the reference point may affect how agents perceive a given level of trust. Changing the default from no trust to full trust leads to an act of commission rather than an act of omission if principals prefer not to entrust the agent with all their money. Many studies in behavioral decision theory find that perceptions of outcomes differ if the outcome resulted from an act of commission rather than from an act omission (for example, Kahneman and Tversky 1982; Baron and Ritov 1994). Thus, the same level of trust may be perceived as less benign when it results from amounts taken (DTG), a commission, rather than from amounts not given (TG), an omission. Such a change in the perception of principals' intentions influences reciprocity, leading to less trustworthiness for any given level of trust in the DTG than in the TG.

[4] Note that an increasing slope is also compatible with certain outcome-based social preference models (see, for example, Ashraf et al. 2004). Thus, we apply further tests to differentiate between the two motives of behavior below.

In addition, while we have focused on trustworthiness levels so far, a change in reference points may also affect the slopes of the trustworthiness curves. Compared with omissions, commissions may induce agents to punish trust withdrawal proportionally more, the more principals take away trust. We refer to this as "trustworthiness responsiveness." It measures how the return ratio changes when trust changes by one increment ($1) and captures the intensity with which agents respond to changes in trust. If differential responses to omissions and commissions drive agents' behavior, we expect a higher degree of trustworthiness responsiveness in the DTG than in the TG.

Reciprocity relies on human intentions and therefore on a human counterpart. Loss aversion would be relevant even if a random mechanism determined the outcome in stage 1. Comparing behavior in our standard games with responses to the random versions of the TG and the DTG allows us to see whether reciprocity has an additional impact on behavior, beyond mere loss aversion.

Trust

Principals may not respond to the different reference points in the DTG and the TG. Our null hypothesis is that trust levels do not differ between the DTG and the TG. In contrast, our behavioral hypothesis is:

Hypothesis 2: Trust levels differ in the DTG from those in the TG.

We offer two conjectures for why we might see either higher or lower trust levels in the TG than in the DTG:

Conjecture 2a: Lower trust levels in the DTG than in the TG may result from principals' accurately predicting how the reference points affect agent behavior and/or from the frames affecting principals' social preferences.

Principals may understand the behavioral effects on agents' trustworthiness of a change in reference points. Since both loss aversion and reciprocity induce less trustworthiness for any given level of trust in the DTG than in the TG, it pays less to trust in the DTG than in the TG. Principals may adjust their behavior accordingly and also trust less in the DTG than in the TG.

In addition, the default in the two games changes principals' decision from *giving* trust to *not taking away* trust. According to Andreoni (1995), giving implies creating a positive externality for the agent while taking imposes a negative externality. If the "warm glow" from giving exceeds the "warm glow" from not taking, principals should offer more trust in the TG than in the DTG.

> **Conjecture 2b:** Higher trust levels in the DTG than in the TG may result from principals' inaccurately predicting how the reference points affect agent behavior: They may neglect the effect of the reference points on average trustworthiness levels and/or on average trustworthiness responsiveness. In addition, principals' own loss aversion could lead to higher trust levels in the DTG than in the TG.

Research on perspective-taking abilities showed that people are often not able to walk in the other party's shoes (for example, Samuelson and Bazerman 1985; Neale and Bazerman 1991). People tend to focus too much on their own thoughts and actions and do not take other parties' actions and motives sufficiently into account (Carroll, Bazerman, and Maury 1988; Simons and Chabris 1999). In binary-choice trust games, for example, principals have been found to fail to adjust accurately to changes in agents' incentives (Bohnet and Huck 2004; Malhotra 2004).

In our design, to predict agent behavior correctly, principals need to form expectations about how the change in the default influences agents' reciprocity and loss aversion. Van Boven et al. (2000; 2003) showed in a series of experiments that people have great difficulty in correctly anticipating the effect of loss aversion on other people. They found that buyers were not able to foresee how the endowment effect influences sellers, and vice versa. Accordingly, our principals may also not foresee how the difference between gains and losses and between omissions and commissions affects their agents' trustworthiness levels and responsiveness.

The change of the default also affects principals' reference points. According to loss aversion and the resulting endowment effect, principals might be more likely to stick with their endowment in the TG than to take the same share in the DTG. Similar effects have been found when framing a game either as a public goods game or a common pool resource game. Whereas in the first game people get an endowment and are asked to *invest* money in a joint project, in

the second game subjects can *take* from a joint project. Although payoffs are equivalent in the two games, people contribute more to the joint project when they have to take their private share from the project than when they have to decide how much to give from their private account to the project (for example, Brewer and Kramer 1986; McCusker and Carnevale 1995).

IV. Results

We first focus on experimentally observed trustworthiness and trust levels in the TG and the DTG and then examine why we might observe differences in behavior based on evidence from the post-experimental questionnaire.

Result 1: Trustworthiness

Trustworthiness levels are lower in the DTG than in the TG.

Return ratios in both games, the TG and the DTG, are below 1. Thus, on average, it does not pay to trust. Table A.1 in the Appendix presents the summary statistics. In the TG, the average return ratio is 0.93. In the DTG, the average return ratio is 0.51. The difference is statistically significant at the 90 percent level using a Mann-Whitney test (p=0.08).[5] However, as trustworthiness might increase with trust, the difference in the observed trustworthiness levels between the DTG and the TG might be due to differences in trust levels. We therefore run regressions where, apart from a dummy variable controlling for being in the DTG (1 if DTG and 0 if TG), we also incorporate as independent variable the amount given/not taken (x).

Table 1 presents the respective ordinary least squares (OLS) regressions, with trustworthiness as the dependent variable. The results in Column (1) show that trustworthiness is positively correlated with trust and that agents return about 60 percentage points less in the DTG than in the TG. Both coefficients are statistically significant at the 95 percent level. Column (2) incorporates an interaction term between trust and the DTG. The slopes in the DTG and the TG do not significantly differ from each other.[6] These results provide preliminary evidence for the absence of differences in trustworthiness responsiveness only. They show trustworthiness

[5] All p-values reported are based on the Mann-Whitney test, unless otherwise noted.

[6] A test of joint significance of the variables *DTG* and *Trust*DTG* shows that their joint association with trustworthiness is statistically significant at the 90 percent level.

conditional on trust actually given/not taken in the TG and the DTG, and not an agent's complete response function, for all possible levels of trust. We examine trustworthiness functions more closely in Result 3, building on the strategy responses in the questionnaire.

Our results support Hypothesis 1: Changing the default option in a trust game changes agents' behavior substantially.

Result 2: Trust

Trust levels are higher in the DTG than in the TG.

Figure 1 presents the distribution of the chosen trust levels (amount given/not taken). Principals do not accurately anticipate the difference in trustworthiness levels between the TG and the DTG. To the contrary, principals are more trusting in the DTG than in the TG. On average, principals send x=\$2.8 to agents in the TG, whereas in the DTG they leave x=\$5.2 with the agent. The difference is statistically significant ($p<0.05$).

Result 2 supports Hypothesis 2. It rejects Conjecture 2a: Principals do not correctly anticipate how the reference point affects agents' behavior, and the warm glow of giving does not outweigh the warm glow of not taking. Principals are substantially more willing to trust agents in the DTG than in the TG.

Looking at the trust and trustworthiness results simultaneously, changing the default from no trust to full trust has a paradoxical effect. The DTG elicits less trustworthiness but more trust than the TG. Consequentially, principals earn less in the DTG than in the TG (\$7.9 vs. \$10.3; $p<0.05$) while agents earn more in the DTG than in the TG (\$23.3 vs. \$15.3; $p<0.01$). Starting in a trust relationship with full trust leads to a more unequal distribution but to a more efficient outcome. Average total earnings by pairs are \$30.3 in the DTG and \$25.6 in the TG ($p<0.05$). Our results show that the default in a trust relationship has substantial behavioral consequences.

Result 3: Lower trustworthiness levels in the DTG than in the TG are mainly due to differences in agents' reciprocal responses to commissions and omissions.

Figure 2 shows the reported return ratios for the trust and the distrust game if either a human or a random mechanism determines the outcomes, based on subjects' responses in the post-experimental questionnaire. We elicited responses to all possible outcomes after stage 1.[7]

We first note that, compatible with our experimental results (Result 1), the level of reported trustworthiness differs substantially between the DTG and the TG. Agents report an average return ratio of 1.04 in the TG and of 0.61 in the DTG ($p<0.01$).[8] The reported trustworthiness levels in the random versions of these games, the RTG and the RDTG, also differ from each other to some degree. The reported average return ratio is 1.09 in the RTG and 0.82 in the RDTG. While this difference is not statistically significant ($p=0.34$), it is still economically meaningful. Thus, we do not want to exclude the possibility that some of the difference in trustworthiness between the TG and the DTG is due to agents' loss aversion (Conjecture 1a).

The random mechanism seems to increase the reported return ratio particularly in the distrust game (but the difference is not significant based on the M-W test, $p=0.33$). There is no difference in reported trustworthiness between the standard and the random versions of the distrust game when no trust is withdrawn. If agents can keep the whole amount, they are willing to return \$10 on average to their principals in both versions of the DTG. Finally, there seems to be a stronger relationship between trust and reported trustworthiness in the regular than the random versions of our games.

To examine these effects more precisely, we run OLS regressions with reported trustworthiness as the dependent variable (Table 2). Column (1) shows estimates of the determinants of reported trustworthiness if the outcome in stage 1 is determined by the principal. Agents reward trust more, the higher a given level of trust. On average, reported trustworthiness is 60 percentage points lower in the DTG than in the TG. The trustworthiness slope in the DTG is somewhat steeper than in the TG. The average trustworthiness responsiveness for all possible levels of trust is 0.05 in the DTG and 0.07 in the TG. This difference (*Given trust levels*DTG* in Column (2)) is not significant. Only for trust levels $x \geq \$5$ do

[7] As trustworthiness (y/x) cannot be computed for $x=0$, we analyze reported trustworthiness for $x>0$.

[8] To examine differences in behavior, we calculate each agent's average reported return ratio for the 10 possible positive trust levels and compare these across the treatments.

the slopes of the reported return ratios in the DTG and the TG significantly differ from each other.[9]

Column (3) shows that if the outcome in the first stage is determined by a random mechanism instead of by the principal, both effects, the correlation between trust and trustworthiness and the correlation between the DTG and trustworthiness, decrease by about half and become insignificant. Agents do not behave reciprocally when confronted with nature and do not punish nature more for "commissions" than for "omissions."

Column (4) combines the two data sets and shows the difference of the differences. Trustworthiness is significantly less correlated with trust levels if a random mechanism determines the outcome in stage 1 than it is in the standard games (*Random*Given Trust Level*), suggesting that agents' behavior is more responsive to the degree of trust when they are confronted with another person rather than nature. Column (4) also shows that agents return larger fractions in the random version of the DTG than in the regular game (*Random*DTG*), suggesting that distrust is perceived as worse when produced by a principal rather than by nature.

Taken together, the higher return ratio in the RDTG than the DTG, the lack of reciprocity and the absence of differential responses to omissions versus commissions in the random versions of our games provide support for Conjecture 1b. Changing the default from no trust to full trust is particularly important when intentions play a role. Controlling for the trust levels, we find that only if the outcome in stage 1 is determined by a person and not by a random mechanism is reported trustworthiness significantly lower in the DTG than in the TG. Agents "punish" principals more severely for committing distrust than for omitting trust, but they hardly respond to nature. Reference-dependent reciprocity is the main driver of behavior.

Result 4: Higher trust levels in the DTG than in the TG are mainly due to principals' expecting no differences in average trustworthiness levels but expecting differences in average trustworthiness

[9] If we run the same regression as in Column (2) for trust levels $x \geq 5$, the coefficient of the interaction *Given trust level*DTG* increases to 0.06 and becomes statistically significant at the 95-percent level.

responsiveness: They expect more trustworthiness for high levels of trust and somewhat less trustworthiness for low levels of trust in the DTG than in the TG.

Result 2 suggests that principals do not take into account accurately the effect of reference points on agents' behavior when deciding how much to trust: Principals are more trusting in the DTG than in the TG although it pays even less (or costs more) to trust in the DTG than in the TG. To further explore why our results reject Conjecture 2a, we look more closely at expectations. Figure 3 shows the expected return ratios for each possible level of trust and how expectations compare with reported return ratios, based on principals' strategy responses in the questionnaire.[10] The figure suggests that, on average, principals expect the same return ratios in the DTG as in the TG but that they adjust to the change in the reference points by expecting trustworthiness to be more responsive to changes in trust levels in the DTG than in the TG.

Table A.1 presents principals' expected average trustworthiness and the respective expected trustworthiness responsiveness. Principals expect an average return ratio of 0.96 in the DTG and 0.90 in the TG (difference not significant). Their expectations are inaccurate in both games. In the DTG, principals are too optimistic about agents' behavior: Agents report that they would only return 0.61 on average ($p<0.05$). In the TG, principals are somewhat too pessimistic about their agents' behavior: Agents report that they would return 1.16 on average ($p=0.17$).

The same pattern holds when comparing expectations with the (experimental) return ratios for the amounts of trust principals chose in the experiment. In the DTG, principals expect a 59 percent premium but only receive back about half of their investment ($E[y/x]=1.59$ as compared with $[y/x]=0.51$, $p<0.01$). In the TG, principals expect back a 36 percent premium on their investment but in fact receive back almost the amount they gave ($E[y/x] =1.36$ as compared with $[y/x]=0.93$, $p=0.17$). The expected return ratios for the trust levels chosen in the experiment do not differ significantly from each other in the two games. On average, principals lack the ability to take the agents' perspective and to anticipate fully the lower trustworthiness in the DTG compared with the TG.

[10] All but two subjects completed the whole questionnaire. For two subjects, one in the TG and one in the DTG, we do not have the whole range of expectations available. They are not part of the analysis of expectations.

At the same time, principals adjust their expectations to the differences in trustworthiness responsiveness in the DTG and the TG. On average, they expect that trustworthiness is more strongly related to trust levels in the DTG than in the TG. Principals expect the return ratio to be reduced by 0.08 for every dollar taken in the DTG, on average. In contrast, in the TG, principals expect that the return ratio will be reduced by only 0.01 for every dollar not given ($p<0.05$). An OLS regression with the expected return ratio as the dependent variable, reported in Table A.2 in the Appendix, suggests that the slopes in the DTG and the TG differ significantly from each other. Principals expect that commissions lead agents to respond more strongly to a change in the level of trust in the DTG than do omissions in the TG. Reported return ratios suggest a similar pattern: Agents respond more negatively to the loss of trust than to the lack of trust, with reported trustworthiness being significantly more responsive to changes in trust in the DTG than in the TG for high trust levels.

In sum, principals correctly anticipate that distrust leads to stronger negative reactions by agents than lack of trust, particularly for high levels of trust, resulting in a closer relationship between trust levels and trustworthiness. But they fail to anticipate that the change in reference points leads to less trustworthiness for all levels of trust, including full trust, in the DTG than in the TG. This partial adjustment of expectations leads to the same average expected trustworthiness in the DTG as in the TG and to higher expected trustworthiness for higher levels of trust and somewhat lower expected trustworthiness for lower levels of trust in the DTG than in the TG.

Table 3 shows that principals' partial adjustment of their expectations relates significantly to their willingness to trust/not distrust. Column (1) shows in an OLS regression with trust (x) as the dependent variable that principals are more trusting in the DTG than in the TG. This effect holds when controlling for average expectations in Column (2). However, when controlling for principals' expectations of the average trustworthiness responsiveness in Column (3), the direct effect of the change in the reference point loses importance and is no longer significant. Trust is related only to expectations of average return ratios and expectations of trustworthiness responsiveness. In Column (4), we use the game frame as an instrument for expectations of trustworthiness responsiveness. Our results suggest that the game frames mainly

affect trust by changing expectations about how responsive trustworthiness is to changes in trust. Subjects expect a stronger reaction to distrust in the DTG than to lack of trust in the TG and thus, are more likely to trust fully.

While Table 3 does not suggest that the reference points exhibit a significant direct influence on behavior, expectations (average and slope) do not account for all of the variance in trust. We focus on the first stage of the TG and the DTG to get a first sense for whether loss aversion could be responsible for the remaining difference. However, in the giving and taking dictator games, changing the default does not affect behavior. When principals were asked about how much they wanted to give to an agent in the triple dictator game, 88 percent decided not to give anything, leaving both parties with their initial endowments of $10. When subjects were asked about how much they wanted to take from an agent, 89 percent took everything, again leaving both parties with $10 ($p=0.86$). The mean amounts given/not taken are basically zero in both games (Table A.1). Thus, we do not find any evidence for loss aversion in the simple dictator games.[11]

Our results are partly in line with Conjecture 2b. Principals entrust the agent with more money in the DTG than in the TG because they do not adjust their expectations of agents' behavior sufficiently. They expect agents to punish commissions more than omissions for low levels of trust. For high levels of trust, they expect higher return ratios in the DTG than in the TG, and this explains the experimentally observed trust levels. Forty percent of the principals choose to entrust their agents with all their money in the DTG, while only 16 percent trust fully in the TG. Our design does not allow us to rule out conclusively that loss aversion may have contributed to our findings. Clearly, the standard dictator games used here did not capture all relevant features of the first-stage trust decision, particularly the risk involved in trusting and the positive amounts that may be returned.

In a related experiment, Bohnet and Meier (2005) employed a "risky dictator game" (Bohnet and Zeckhauser 2004) to examine loss aversion in the face of risk. They confronted

[11] Generally, there seems to be a surprising willingness to take away or "steal" money from others in experiments, see, for example, Eichenberger and Oberholzer-Gee (1998) and Falk and Fischbacher (2002).

principals with a binary choice: either a sure outcome of ($10, $10) for the principal and the agent or a gamble with 1/3 chance of ($20/$20), a 1/3 chance of ($10, $30), that is, $10 for the principal and $30 for the agent, and a 1/3 chance of ($0, $40), that is, nothing for the principal and $40 for the agent. Principals started with either ($10, $10) in the TG condition or ($0, $40) in the DTG condition. They could either keep their $10 (take $30 in the DTG condition) or give their $10 to the agent (take $0 from the agent in the DTG condition). If they gave everything (did not take anything), they ended up with the gamble. In the TG condition of the risky dictator game, 42 percent decided to give everything; in the DTG condition, 69 percent decided not to take anything. The difference is statistically significant (p=0.05). In line with loss aversion, subjects were more likely to risk their sure share when their starting point was $0 than when it was $10.

V. Conclusions

People do not distrust enough. In our distrust games, where the default is full trust, principals lose money by continuing to trust. They are about 20 percent worse off than if they had withdrawn all trust from their agents. Their agents, in contrast, are substantially better off than if they were not trusted. They experience an increase in their earnings of 130 percent compared with the no-trust equilibrium. Clearly, the magnitude of these numbers is closely related to our experimental set-up. However, the substantive inequality created reminds one of what many workers and investors experienced after the implosion of some of the big corporations.

We compare the distrust game with a standard trust game, in which the default is no trust. Agents reward trust less in the distrust game, where trust is the status quo, than in the trust game, where trust corresponds to an active leap of faith. This is mainly due to reciprocal preferences, which punish loss of trust, a commission, more heavily than lack of trust, an omission. As a consequence, principals earn more in the trust game than in the distrust game. In the trust game, they earn slightly more than they would have without any trust (and agents earn about 50 percent more). Yet, principals trust less in the trust game than in the distrust game, where they lose money on average.

Why do principals not anticipate this? Our data suggest that they are not able to take their agent's perspective accurately. They correctly expect that trustworthiness responds more strongly to changes in the level of trust in the distrust game than in the trust game. A return ratio that strongly decreases with the amount of distrust makes it more attractive to withdraw little trust or nothing at all. Yet, agents also reward full trust less in the distrust game than in the trust game. For them, full trust in the former just conforms to the status quo, while it represents a gift in the latter. Principals neglect to see that a change in reference point from no trust to full trust affects agents' responses for *all* possible levels of trust.

Thus, framing matters. Our results are in line with behavioral decision-theoretic models assuming reference-point dependent preferences. Changing the reference point from no trust to full trust does not change incentives but has substantial behavioral consequences. To what degree our results can account for the lack of distrust observed outside of the laboratory is obviously an open question. Did Enron or WorldCom's managers feel as little indebted to their shareholders and stakeholders as our agents in the experiment? Even when our agents were trusted fully in the distrust game, they did not share the benefits of this investment with their principals. With full trust, principals expected to make money but, on average, lost money.

And were the shareholders and stakeholders as optimistic as our principals? Bazerman and Watkins (2004) describe the U.S. government's and the private sector's optimism in the late 1990s despite repeated warnings by the SEC and a number of academics that the lack of auditor independence was destined to lead to disaster. Clearly, some of this optimism was due to misaligned incentives, with some of the key decision-makers receiving large amounts of money from these industries. Our results suggest that even if we hold incentives constant, a default of full trust leads principals to wrongly assess their agents' trustworthiness and, as a consequence, to end up trusting too much.

Research on other principal-agent relationships, namely, between patients and doctors, suggests that principals generally trust their agents. Cain, Loewenstein, and Moore (2005) report that "people tend to be naturally trusting and credulous toward their own advisors. In the domain of medicine, for example, research shows that while many people are ready to

acknowledge that doctors generally might be affected by conflicts of interest, few can imagine that their own doctors would be affected (Gibbons, Landry, Blouch, Jones, Williams, Lucey, and Kroenke 1998)." Patients only rarely seek second opinions (for example, Foreman 2001).

In our experiments, the default of full trust increased efficiency and inequality, with principals much worse off than agents and than they would have been if they had withdrawn all trust. In the long run, behavior away from the equilibrium path where principals keep losing out is not sustainable. If we want to avoid a fallback to complete distrust, widely acknowledged in the popular press to have occurred after the recent scandals (for exmaple, Callahan, 2004; Lorsch, Berlowitz, and Zelleke, 2005), and preserve some of the efficiency gains that go along with trust, a default of no trust is more likely to yield success. Trust should not be taken for granted. Rather, second opinions should be the standard procedure in doctor-patient relationships and parents should be encouraged to pay surprise visits to their children's schools. Clearly, in an ideal world, we would align agents' and principals' incentives and avoid the conflicts of interest inherent in many trust relationships. The Sarbanes-Oxley Act tries to deter agents from doing wrong but is unlikely to be successful until true auditor independence is guaranteed and auditors' and shareholders' interests are more closely aligned.

References

Andreoni, James. 1995. "Warm-Glow Versus Cold-Prickle: The Effects of Positive and Negative Framing in Cooperation in Experiments." *Quarterly Journal of Economics* 110(1): 1–21.

Andreoni, James, and John H. Miller. 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica* 70(2): 737–53.

Ashraf, Nava, Iris Bohnet, and Nikita Piankov. 2004. "Decomposing Trust and Trustworthiness." *Mimeo*, Kennedy School of Government, Harvard University.

Baron, Jonathan and Ilana Ritov. 1994. "Reference Points and Omission Bias." *Organizational Behavior and Human Decision Processes* 59:475–98.

Barr, Abigail. 2003. "Trust and Expected Trustworthiness: Experimental Evidence from Zimbabwean Villages." *Economic Journal* 113(489): 614–30.

Bazerman, Max H., and Michael D. Watkins. 2004. *Predictable Surprises: The Disasters You Should Have Seen Coming, and How to Prevent Them*. Cambridge: Harvard Business School Press.

Bazerman, Max H., Kimberly P. Morgan, and George F. Loewenstein. 1997. "The Impossibility of Auditor Independence." *Sloan Management Review* 38(4): 89–94.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10(1): 122–42.

Bohnet, Iris, and Steffen Huck. 2004. "Repetition and Reputation: Implications for Trust and Trustworthiness When Institutions Change." *American Economic Review* 94(2): 362–66.

Bohnet, Iris, and Stephan Meier. 2005. "Taking Risk in Negotiations." *Mimeo*, Kennedy School of Government, Harvard University.

Bohnet, Iris, and Richard Zeckhauser. 2004. "Trust, Risk, and Betrayal." *Journal of Economic Behavior & Organization* 55(4): 467–84.

Bolton, Gary, and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity and Competition." *American Economic Review* 90(1): 166–93.

Brewer, Marilynn B., and Roderick M. Kramer. 1986. "Choice Behavior in Social Dilemmas: Effects of Social Identity, Group Size, and Decision Framing." *Journal of Personality and Social Psychology* 50(3): 543–49.

Cain, Daylian M., George Loewenstein, and Don A. Moore. 2005. "The Dirt on Coming Clean: Perverse Effects of Disclosing Conflicts of Interest." *Journal of Legal Studies* 34(1): 1–25.

Callahan, David. 2004. *The Cheating Culture: Why More Americans Are Doing Wrong to Get Ahead*. Orlando: Harcout, Inc.

Camerer, Colin. 2003. *Behavioral Game Theory*. Princeton: Princeton University Press.

Camerer, Colin, and Keith Weigelt. 1988. "Experimental Tests of a Sequential Equilibrium Reputation Model." *Econometrica* 56(1): 1–36.

Cardenas, Juan, and Jeffrey Carpenter. 2005. "Experimental Development Economics: A Review of the Literature and Ideas for Future Research." *Mimeo,* Middlebury College.

Carroll, J.S., Max H. Bazerman, and R. Maury. 1988. "Negotiator Cognitions: A Descriptive Approach to Negotiators' Understanding of Their Opponents." *Organizational Behavior and Human Decision Processes* 41:352–70.

Cox, James C. 2004. "How to Identify Trust and Reciprocity." *Games and Economic Behavior* 46(2): 260–281.

Croson, Rachel T.A. 2000. "Thinking Like a Game Theorist: Factors Affecting the Frequency of Equilibrium Play." *Journal of Economic Behavior and Organization* 41:299–314.

Economist, The. 2005. "Attempts to Reform Accounts Are Creating Their Own Problems." July 28: 65.

Eichenberger, Reiner, and Felix Oberholzer–Gee. 1998. "Rational Moralists. The Role of Fairness in Democratic Economic Policy." *Public Choice* 94:191–210.

Ensminger, J. 2000. "Experimental Economics in the Bush: Why Institutions Matter." In *Institutions, Contracts, and Organizations: Perspectives from New Institutional Economics*, ed. C. Menard, 158–171. Cheltenham: Edward Elgar Publishing.

Falk, Armin, and Urs Fischbacher. 2002. "'Crime' in the Lab—Detecting Social Interaction." *European Economic Review* 46(4–5): 859–869.

Fehr, Ernst, and Klaus Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114(3): 817–868.

Fehr, Ernst, and Klaus Schmidt. 2002. "Theories of Fairness and Reciprocity – Evidence and Economic Application." In *Advances in Economics and Econometrics—8th World Congress, Econometric Society Monographs*, eds. M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, 208–257. Cambridge: Cambridge University Press.

Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl. 1993. "Does Fairness Prevent Market Clearing? An Experimental Investigation." *Quarterly Journal of Economics* 108(2): 437–459.

Foreman, Judy. 2001. "First Rule: Don't Hesitate to Get Second Opinion." http://www.myhealthsense.com/F010522_secondOpinion.html.

Gibbons, Robert, Frank J. Landry, Denise L. Blouch, David L. Jones, Frederick K. Williams, Catherine R. Lucey, and Kurt Kroenke. 1998. "A Comparison of Physicians' and Patients' Attitudes toward Pharmaceutical Industry Gifts." *Journal of General Internal Medicine* 13:151–54.

Greig, Fiona, and Iris Bohnet. 2005. "Is There Reciprocity in a Reciprocal–Exchange Economy? Evidence from a Nairobi Slum." *Mimeo*, Kennedy School of Government, Harvard University.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1986. " Fairness and the Assumptions of Economics." *Journal of Business* 59(4) Part 2 October: S285–S300.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47(2): 263–291.

Kahneman, Daniel, and Amos Tversky. 1982. "The Psychology of Preferences." *Scientific American* 246:160–73.

Kreps, David M. 1990. "Corporate Culture and Economic Theory." In *Perspectives on Positive Political Economy*, eds. J. E. Alt and K. A. Shepsle, 90–143. Cambridge: Cambridge University Press.

Lorsch, Jay W., Leslie Berlowitz, and Andy Zelleke, eds. 2005. *Restoring Trust in American Business*. Cambridge: The MIT Press.

Malhotra, Deepak. 2004. "Trust and Reciprocity Decisions: The Differing Perspectives of Trustors and Trusted Parties." *Organizational Behavior and Human Decision Processes* 94(2): 61–73.

McCabe, Kevin, Mary Rigdon, and Vernon Smith 2003. "Positive Reciprocity and Intentions in Trust Games." *Journal of Economic Behavior and Organization* 52(2): 267–275.

McCusker, Christopher, and Peter J. Carnevale. 1995. "Framing in Resource Dilemmas: Loss Aversion and the Moderating Effects of Sanctions." *Organizational Behavior and Human Decision Processes* 61(2): 190–201.

Meier, Stephan. 2004. "A Survey on Economic Theories and Empirics on Pro–Social Behavior." *Mimeo,* Institute for Empirical Research in Economics, University of Zurich.

Neale, Margaret A., and Max H. Bazerman. 1991. *Cognition and Rationality in Negotiation*. New York: The Free Press.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review* 83(5): 1281–1302.

Samuelson, William, and Max H. Bazerman. 1985. "Negotiating under the Winner's Curse." In *Research in Experimental Economics*, ed. V. Smith, 105-137. Greenwich: JAI Press.

Samuelson, William, and Richard Zeckhauser. 1988. "Status Quo Bias in Decision Making." *Journal of Risk and Uncertainty* 1(1): 1–53.

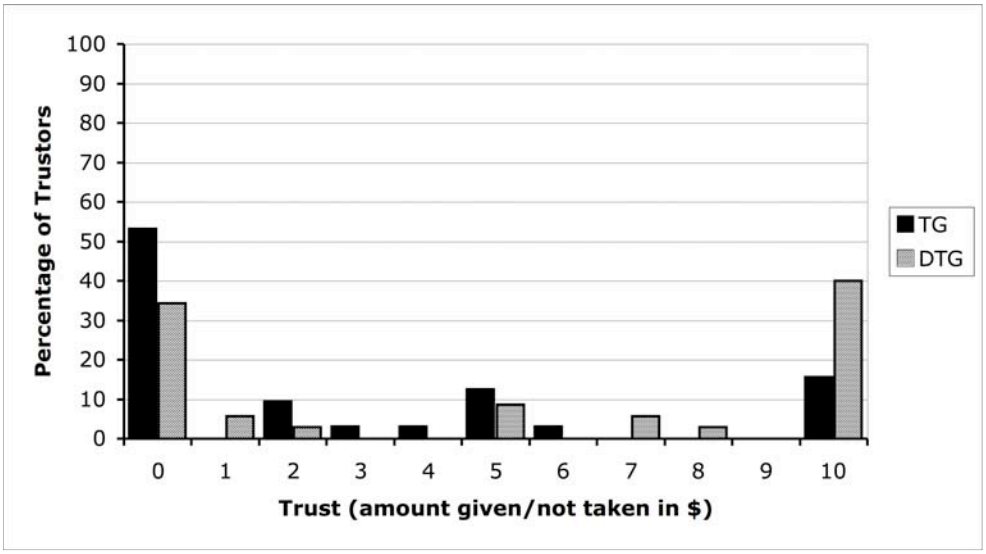
Simons, D.J., and C.F. Chabris. 1999. "Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events." *Perception* 28:1059–1074.

Thaler, Richard H. 1980. "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization* 1 (March): 39–60.

van Boven, Leaf, David Dunning, and George Loewenstein. 2000. "Egocentric Empathy Gaps between Owners and Buyers: Misperceptions of the Endowment Effect." *Journal of Personality and Social Psychology* 79(1): 66–76.

van Boven, Leaf, David Dunning, and George Loewenstein. 2003. "Mispredicting the Endowment Effect: Underestimation of Owners' Selling Prices by Buyer's Agents." *Journal of Economic Behavior and Organization* 51(3): 351–65.

Figure 1: Trust: Distribution of Amounts Given by Principals



Notes: Trust game: Trust=x. Distrust game: Trust=amount not taken: x=10-(z/3).

Figure 2: Trustworthiness: Reported Return Ratios

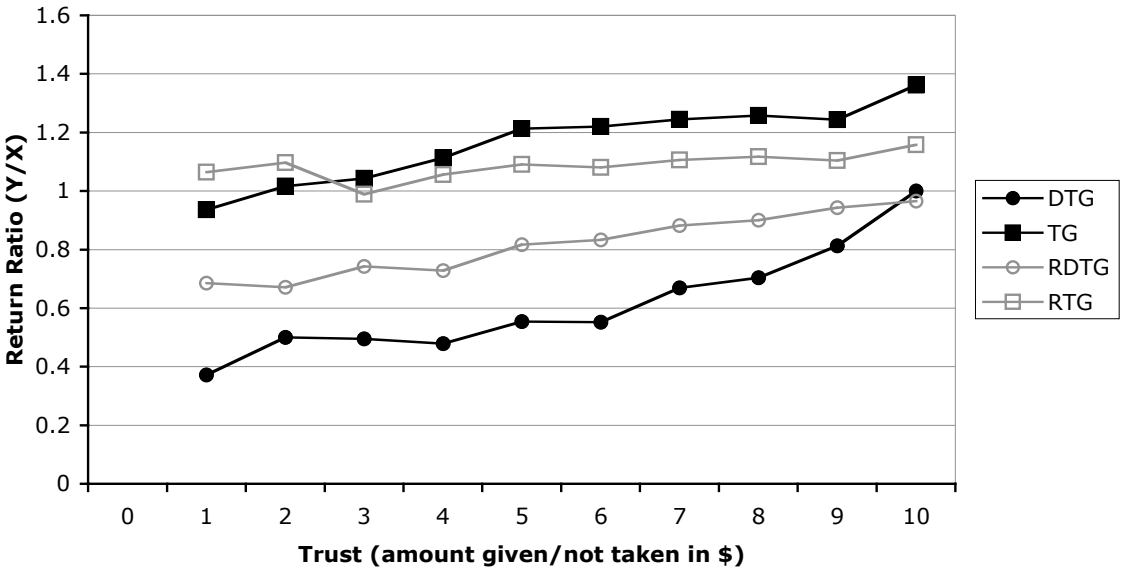


Figure 3: Expected and Reported Return Ratios

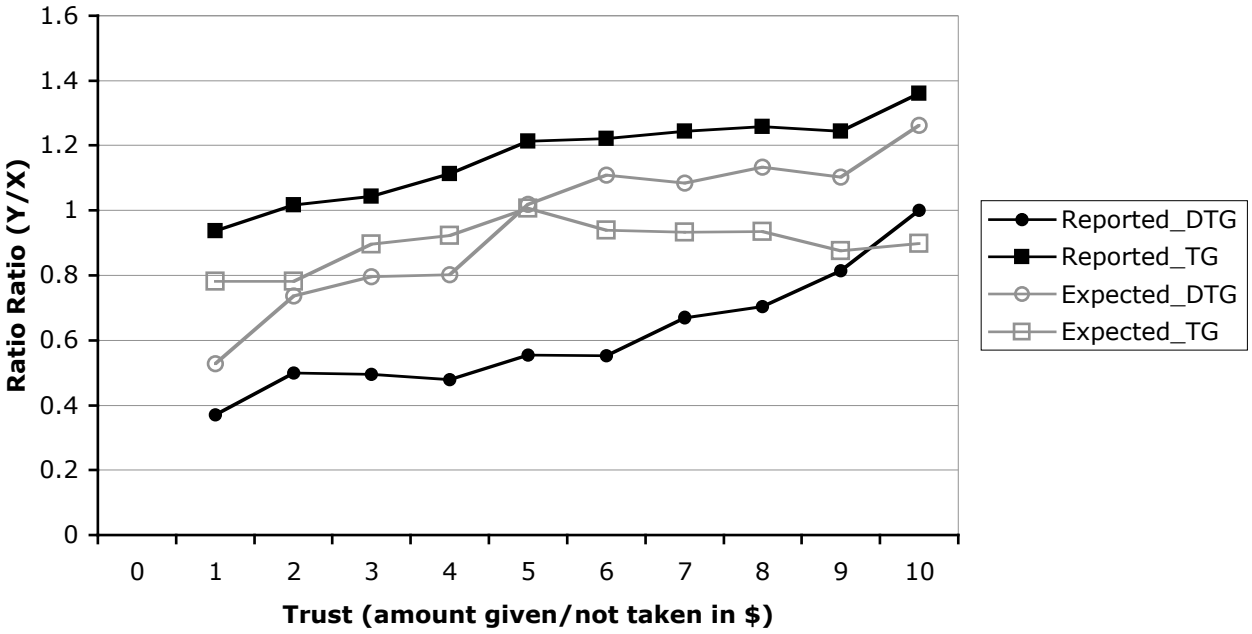


Table 1: Determinants of Trustworthiness

	(1)	(2)
Trust (amount given/not taken)	0.078*	0.096
	(0.031)	(0.063)
DTG (=1)	-0.573*	-0.373
	(0.254)	(0.419)
Trust*DTG		-0.030
		(0.070)
Constant	0.470(*)	0.363
	(0.241)	(0.366)
R-squared	0.165	0.168
# of observations	38	38

Notes: Dependent variable: trustworthiness (*y/x*). OLS-Regressions with robust standard errors in parentheses. F-test for joint significance of *DTG* and *Trust*DTG*: F (2,34)=2.71; p>0.081.

(*) Significant at 10 percent.

* Significant at 5 percent.

** Significant at 1 percent.

Table 2: Reported Trustworthiness (Return Ratio) after Decision by Person or Random Mechanism

	(1) with Person	(2) with Person	(3) Random	(4) All
Given Trust Level	0.050**	0.042*	0.023	0.050**
	(0.013)	(0.018)	(0.017)	(0.013)
DTG (=1)	-0.551**	-0.636*	-0.269	-0.551**
	(0.186)	(0.264)	(0.214)	(0.186)
Given Trust Level*DTG		0.015		
		(0.026)		
Random Allocation (=1)				0.070
				(0.150)
Random*Given Trust Level				-0.027*
				(0.011)
Random*DTG				0.282(*)
				(0.142)
Constant	0.890**	0.935**	0.960**	0.890**
	(0.178)	(0.204)	(0.211)	(0.178)
R-squared	0.1203	0.1209	0.0228	0.0679
# of observations	660	660	660	1320
# of individuals	66	66	66	66

Notes: Dependent variable: Reported trustworthiness. Robust standard errors in parentheses. Clustered for individuals.

(*) Significant at 10 percent.

* Significant at 5 percent.

** Significant at 1 percent.

Table 3: Determinants of Trust (Amount given/not taken)

	(1)	(2)	(3)	(4)
DTG (=1)	2.248*	2.130*	0.861	
	(1.019)	(0.968)	(0.882)	
Expected Return Ratio		1.966**	2.256**	2.453**
		(.702)	(0.563)	(0.680)
Expected Trustworthiness Responsiveness			18.255**	30.650*
			(2.971)	(12.713)
Constant	2.781**	1.019	0.525	0.189
	(0.655)	(0.665)	(0.585)	(0.849)
R-squared	0.0699	0.191	0.431	0.320
# of observations	66	66	66	66

Notes: Dependent variable: trust (*x*). Models (1)-(3) show OLS-regressions with robust standard errors in parentheses. Model (4) shows a 2SLS regression with robust standard errors in parentheses. First-stage regression: Expected Association of Trust and Trustworthiness = 0.069* (s.e. = 0.029) DTG - 0.0159 (0.019) Average Expected Return Ratio + 0.027 (0.027) constant. Adj. R-squared: 0.063.

(*) Significant at 10 percent.

* Significant at 5 percent.

** Significant at 1 percent.

Table A.1: Summary statistics

	Trust-Game (TG)	Distrust-Game (DTG)	Mann-Whitney Test
Trustworthiness (return ratio) by agent (y/x)	0.933	0.51	$\|z\| = 1.747$
	(0.821)	(0.746)	$p < 0.1$
	[15]	[23]	
Amount returned by agent (y)	3.063	3.057	$\|z\| = 0.520$
	(5.842)	(6.058)	$p = 0.603$
	[32]	[35]	
Trust (amount given/not taken) by principal (x)	2.781	5.171	$\|z\| = 2.130$
	(3.705)	(4.560)	$p < 0.05$
	[32]	[35]	
Earnings by agents	15.281	23.314	$\|z\| = 2.747$
	(7.646)	(12.153)	$p < 0.01$
	[32]	[35]	
Earnings by principals	10.281	7.886	$\|z\| = 2.251$
	(3.846)	(5.556)	$p < 0.05$
	[32]	[35]	
Efficiency (earnings by pairs)	25.563	30.343	$\|z\| = 2.130$
	(7.409)	(9.120)	$p < 0.05$
	[32]	[35]	
Questionnaire			
Reported trustworthiness (with human principal)	1.165	0.614	$\|z\| = 2.982$
	(0.151)	(0.112)	$p < 0.01$
	[31]	[35]	
Reported trustworthiness (with random mechanism)	1.086	0.817	$\|z\| = 0.950$
	(0.171)	(0.130)	$p = 0.342$
	[31]	[35]	
Reported trustworthiness responsiveness (slope: $\partial(y/x)/\partial x$)	0.047	0.070	$\|z\| = 0.828$
	(0.121)	(0.120)	$p = 0.408$
	[31]	[35]	
Expected trustworthiness by principal	0.896	0.957	$\|z\| = 0.485$
	(0.147)	(0.119)	$p = 0.628$
	[32]	[34]	
Expected trustworthiness for chosen amount of Trust	1.356	1.589	$\|z\| = 0.519$
	(0.796)	(0.521)	$p = 0.604$
	[15]	[23]	
Expected trustworthiness responsiveness	0.013	0.081	$\|z\| = 2.110$
	(0.120)	(0.113)	$p < 0.05$
	[32]	[34]	
Amount given/not taken in Dictator Game	0.438	0.471	$\|z\| = 0.136$
	(0.220)	(0.308)	$p = 0.892$
	[32]	[34]	

Notes: We report means. Standard deviation in parentheses. Number of observations in brackets.

Table A.2: Determinants of Expected Trustworthiness

	(1) TG	(2) DTG	(3) All
Given Trust Level	0.011	0.071**	0.011
	(0.018)	(0.019)	(0.018)
DTG (=1)			-0.271
			(0.248)
DTG*Given Trust Level			0.060*
			(0.026)
Constant	0.835**	0.564**	0.835**
	(0.195)	(0.157)	(0.193)
Adj. R-squared	0.001	0.064	0.032
# of observations	320	340	660
# of individuals	32	34	66

Notes: OLS-regressions with robust standard errors in parentheses. Clustered for individuals. Column (1) for trust game (TG), Column (2) for distrust game (DTG), and Column (3) for all together.

(*) Significant at 10 percent.

* Significant at 5 percent.

** Significant at 1 percent.

Instructions

Welcome to our research project!

How this study is conducted. This study is conducted anonymously. Participants will be identified only by code numbers. The instructions are the same for all the participants and are self-explanatory. If you have any questions after we have read the instructions aloud, please raise your hand and someone will come by to assist you. *We ask that you do not talk during the experiment.*
In this study, half the participants are randomly assigned to be Person X, the other half to be Person Y. A Person X is paired with a Person Y. Both are present in this room. You are not told who this person is either during or after the experiment nor is s/he told who you are. All participants in the experiment are currently reading the same set of instructions.

You are randomly chosen to be **Person X**.

[NEXT PAGE]

ABOUT THE DECISION

The study is conducted in two stages:

Stage 1
Each Person X is allocated $0; Person Y $40. Person X makes the first decision. Person X can decide how much out of $30 Person Y currently holds, Person X wants to take. Each dollar taken by Person X is divided by three before Person X receives it. For example, if Person X takes $3 from Person Y, Person X receives $1.
Person X can take any of the following amounts (including zero):
$30, $27, $24, $21, $18, $15, $12, $9, $6, $3, $0.

Stage 2
Person Y then decides how much of the amount s/he holds after Stage 1 to give to Person X. Person X will receive exactly the amount of money given by Person Y. For example, if Person Y gives $3, Person X receives $3.
Person Y can give any amount, in whole dollars, equal to or smaller than the amount of money s/he holds after Stage 1 (including zero).

Example:
In Stage 1, if a Person X decides to take $12 from Person Y:
Person X' payoffs after Stage 1 are $12/3=$4;
Person Y's payoffs after Stage 1 are $40-12=$28.
In Stage 2, Person Y can give any amount of money out of his/her $28 to Person X. For example, if Person Y decides to give $7:
Person X's payoffs after Stage 2 are $4+$7= $11;
Person Y's earning are $28-$7=$21.

[NEXT PAGE]

THE DETAILS OF THE EXPERIMENT

Conduct of study:
(i) You randomly receive a unique code number. You are randomly chosen to be **Person X**.
(ii) Person X and Person Y receive the quiz.
(iii) All answer the questions in the quiz. We collect these forms. If you are Person X, you receive an unmarked envelope with a decision form.
(iv) We start with Stage 1. Persons X decide how much to take from Persons Y. After having indicated their decisions on the decision form, Persons X put their decision form back into the unmarked envelope.
(v) Decision sheets in the unmarked envelopes are randomly distributed to Persons Y.
(vi) We continue with Stage 2. Persons Y decide how much of the money they hold after Stage 1 to give to their Person X. After having indicated their decisions on the decision form, Persons Y put the decision form back in the unmarked envelope.
(vii) Person Y knows the final outcome; Person X is informed on the outcome when collecting his/her earnings.
(viii) We calculate your earnings.

Completion of the study and earnings:
- After the study, we ask you to complete a questionnaire.
- You can collect your earnings by presenting your CODE NUMBER FORM at the end of the study. Person X can check the original decision sheet. Your earnings will be in an envelope only marked with your code number.

[NEXT PAGE]

QUIZ

Before you make your decisions, please answer the following questions.
When you have completed the quiz, please raise your hand.

Example 1: **Person X takes $15 from Person Y in Stage 1.**
What are the payoffs of Person X and Person Y after Stage 1:

X's payoffs after Stage 1:
Y's payoffs after Stage 1:
Person Y gives $15 in Stage 2.
What are the final payoffs of Person X and Person Y after Stage 2:
X's payoffs after Stage 2:
Y's payoffs after Stage 2:

Example 2: **Person X takes $0 from Person Y in Stage 1.**
What are the payoffs of Person X and Person Y after Stage 1:
X's payoffs after Stage 1:
Y's payoffs after Stage 1:
Person Y gives $15 in Stage 2.
What are the final payoffs of Person X and Person Y after Stage 2:
X's payoffs after Stage 2:
Y's payoffs after Stage 2:

Example 3: **Person X takes $15 from Person Y in Stage 1.**
What are the payoffs of Person X and Person Y after Stage 1:
X's payoffs after Stage 1:
Y's payoffs after Stage 1:
Person Y gives $0 in Stage 2.
What are the final payoffs of Person X and Person Y after Stage 2:
X's payoffs after Stage 2:
Y's payoffs after Stage 2:

[NEXT PAGE]

DECISION FORM
[Please do not forget to indicate your code number]

Stage 1: **Decision of Person X**
Your code number is: _____
As Person X, how much money (if any) do you take from Person Y?
Please check one:

I take:			X's payoffs after Stage 1:	Y's payoffs after Stage 1:
❒	$30	→	10	10
❒	$27	→	9	13
❒	$24	→	8	16
❒	$21	→	7	19
❒	$18	→	6	22
❒	$15	→	5	25
❒	$12	→	4	28
❒	$9	→	3	31
❒	$6	→	2	34
❒	$3	→	1	37
❒	$0	→	0	40

Stage 2: **Decision of Person Y**
Your code number is: _____
As Person Y, how much money (if any) do you give to Person X from the amount of money you hold after Stage 1?

I give:		X's payoffs after Stage 2:	Y's payoffs after Stage 2:
$ _____	→	$ _____	$ _____

X's payoffs after Stage 1:
Y's payoffs after Stage 1:
Person Y gives $0 in Stage 2.
What are the final payoffs of Person X and Person Y after Stage 2:
X's payoffs after Stage 2:
Y's payoffs after Stage 2:

[NEXT PAGE]

DECISION FORM
[Please do not forget to indicate your code number]

Stage 1: **Decision of Person X**

Your code number is: _____

As Person X, how much money (if any) do you take from Person Y?

Please check one:

I take:			X's payoffs after Stage 1:	Y's payoffs after Stage 1:
❒	$30	→	10	10
❒	$27	→	9	13
❒	$24	→	8	16
❒	$21	→	7	19
❒	$18	→	6	22
❒	$15	→	5	25
❒	$12	→	4	28
❒	$9	→	3	31
❒	$6	→	2	34
❒	$3	→	1	37
❒	$0	→	0	40

Stage 2: **Decision of Person Y**

Your code number is: _____

As Person Y, how much money (if any) do you give to Person X from the amount of money you hold after Stage 1?

I give:		X's payoffs after Stage 2:	Y's payoffs after Stage 2:
$ _____	→	$ _____	$ _____