Deciding to Distrust

Iris Bohnet and Stephan Meier

August 2005

RWP05-049

Deciding to Distrust

IRIS BOHNET[a] AND STEPHAN MEIER[b]

August 18, 2005

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. Holding the expected returns from trusting constant, people are more trusting when the default is trust than when it is no trust. In a new game, the Distrust Game (DTG), where the default is full trust, trust levels are higher and trustworthiness levels lower than in the BDM-Trust Game (TG), where the default is no trust. Agents punish distrust more in the DTG than in the TG but principals do not anticipate this.

Keywords: Trust, reciprocity, framing, experiments

[a] Kennedy School of Government, Harvard University, 79 JFK Street, Cambridge MA 02138, USA; email: iris_bohnet@harvard.edu.
[b] Kennedy School of Government, Harvard University, 79 JFK Street, Cambridge MA 02138, USA; email: stephan_meier@ksg.harvard.edu.
We are grateful for helpful comments from Theresa Lund. The first author gratefully acknowledges financial support from the Dean's Research Fund at the Kennedy School of Government and the Program on Negotiation at Harvard Law School, and the second author from the Swiss National Science Foundation. This paper was written while the second author was a research fellow with the Women and Public Policy Program at the Kennedy School of Government. He thanks WAPPP for its hospitality.

The recent scandals in big corporations pose many questions. Among them is the puzzle 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. Holding the expected returns (costs) from trusting constant, 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 to a reference point of no trust, principals will not use as much scrutiny and 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 with a certain amount of money and then decides with how much she[2] wants to entrust 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

[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 (for a discussion of different approaches, see Bohnet and Croson, 2004).

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

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 (e.g., 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 preferences models do not expect the default to affect 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 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 (e.g., 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 than in the trust game for given trust levels because of loss aversion (Kahneman and Tversky, 1979) and reciprocal preferences (Rabin, 1993), which punish commissions more heavily than omissions.

For principals, the story is more complicated. If principals have rational expectations of agent behavior, understanding that the expected returns from trusting (i.e., giving in the TG and not taking in the DTG) are smaller in the distrust than 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 agent's shoes, their expectations may not be accurate and lead to similar or even higher trust levels in the distrust than in the trust game. Psychological research on perspective taking suggests that an unbiased perception of other people's preferences and beliefs is hard. For example, 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 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 commissions harder than omissions. However, principals do not anticipate this. They do not foresee that agents in the distrust game are much less trustworthy than in the trust game. They are too optimistic about returns and trust their agents even more in the distrust game than in the trust game. Thus, principals respond to their—albeit inaccurate—expectations of trustworthiness.

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, only differ in the starting point, no trust or full trust, but not in their payoff space and incentives.

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 agent in the TG or how much to take from their agent 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 out of their payoffs after stage 1 they wanted to return to their Persons X.

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, that is, other-regarding concerns[3] and loss aversion (Kahneman and Tversky, 1979), and/or intention-based preferences, i.e., 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 agent 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 like 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)

[3] For reviews of theories and empirical evidence on social preferences, see Fehr and Schmidt (2003) and Meier (2004).

only in that the first decision is determined by a random mechanism. Agents' 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, again using the strategy method, for each possible level of trust. To get a sense for 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 was 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 (Kahneman, Knetsch and Thaler, 1986; 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 in private at the end of a session. On average, subjects earned \$24 (including the show-up fee of \$10) for a 30-40 minute experiment.

III. Behavioral Hypotheses

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 only care 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 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 non-trivial 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 Greig and Bohnet, 2005; Cardenas and Carpenter, 2005) to about 50 percent of the endowment in developed countries (for an overview, see Camerer, 2003). Trustworthiness varies between 54 percent of the amount given for Kenya (Ensminger, 2000) and 128 percent of the amount given in Zimbabwe (Barr, 2003). In developed countries, agents return on average the amount given, i.e., 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 (e.g., 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 do not only care about the distribution of material payoffs but also about how the distribution evolved, i.e., what the intentions of the other party were.

Both aspects of social preferences might be sensitive to framing effects: 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 frames 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 are lower in the DTG than in the TG:

$[y/x]_{DTG} < [y/x]_{TG}$.

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 than in the gain domain. They assign more weight to potential losses than potential gains (Kahneman and Tversky, 1979). Due to loss aversion, people favor the status quo to 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 to 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 should be lower in the DTG than in the TG.

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

Agents' trustworthiness is generally 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, that is $\partial(y/x)/\partial x>0$.[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 a commission rather than an omission if principals prefer not to entrust the agent with all their money. A large number of studies in behavioral decision theory finds that perceptions of outcomes differ if the outcome resulted from commission rather than from omission (e.g., 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 on average in the DTG than in the TG, that is $[y/x]_{DTG}<[y/x]_{TG}$. Compared to

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

omissions, commissions may lead to a fixed proportional decrease in trustworthiness for each level of trust.[5]

Reciprocity relies on human intentions and therefore on a human counterpart. Loss aversion is 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, above loss aversion.

Trust

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

Hypothesis 2: Trust levels differ in the DTG and the TG: $x_{DTG} \neq x_{TG}$.

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

Conjecture 2a: Higher trust levels in the TG than the DTG may result from principals accurately expecting how the frames affect agent behavior or from the frames affecting principals' social preferences.

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

In addition, the default in the two games changes principals' decision from *giving* trust to not *taking* away trust. According to Andreoni (1992), giving implies creating a positive externality while taking imposes a negative externality on the agent. If the "warm glow" from

[5] In addition, the commission/omission effect may also affect the slopes of the trustworthiness curves. Compared to omissions, commissions may induce agents to punish trust withdrawal proportionally more, the more principals take away trust. We could think of this as a commission cost that is constant in absolute terms for all levels of trust, leading to $[\partial(y/x)/\partial x]_{DTG} > [\partial(y/x)/\partial x]_{TG}$. We focus on level effects here and base our predictions on a fixed proportional commission cost for all levels of trust but control for the relevance of slope effects empirically.

giving exceeds the "warm glow" from not taking, principals should offer more trust in the TG than in the DTG.

> **Conjecture 2b:** Lower trust levels in the TG than the DTG may result from principals wrongly expecting how the frames affect agent behavior or from principals' own loss aversion.

Research on perspective taking abilities showed that people are often not able to walk in the other party's shoes (e.g. 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 (Simons and Chabris, 1999; Carroll, Bazerman and Maury, 1988). In binary-choice trust games, for example, principals have been found to not accurately adjust to changes in agents' incentives (Bohnet and Huck, 2004; Malhotra, 2004).

In our design, to correctly predict agent behavior, 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 difficulties to correctly anticipate 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.

The change of the default also affects principals' reference points. According to loss aversion and the resulting endowment effect, principals should 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 (e.g. 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.

Figure 1 presents the distribution of the trustworthiness levels graphically. The figure shows that agents' trustworthiness, the return ratio for positive amounts sent by principals, is much lower in the DTG than in the TG. In the TG, the return ratio is 0.93 on average. In the DTG, the average return ratio is 0.51, which is almost half the trustworthiness measured in the TG. The difference is statistically significant at the 90 percent level using a Mann-Whitney test ($p=0.08$).[6] Table A.1 in the Appendix presents the summary statistics. Return ratios in both games, the TG and the DTG, are below 1. Thus, on average, it does not pay to trust.

[Figure 1 about here]

Table 1 presents an OLS-regression with trustworthiness as the dependent variable. Apart from a variable controlling for being in the DTG, the regressions incorporate as independent variable also the amount given/not taken (x). The results in Column (1) show that trustworthiness is positively correlated with trust and that agents return 60 percentage points less in the DTG than in the TG. Column (2) incorporates an interaction term between trust and DTG. The slopes in the DTG and the TG do not significantly differ from each other. Clearly, these results provide preliminary evidence on the trustworthiness functions only. They show trustworthiness *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: $[y/x]_{DTG}<[y/x]_{TG}$

[Table 1 about here]

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

Result 2: Trust

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

Figure 2 presents the distribution of the chosen trust levels (amount given/not taken). Principals do not 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 ($p<0.05$).

[Figure 2 about here]

Result 2 supports Hypothesis 2: $x_{DTG} \neq x_{TG}$. It rejects Conjecture 2a: Principals do not anticipate how framing 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 the paradoxical effect that 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 3 shows the reported return ratios for the trust and the distrust game if either a human or a random mechanism determined the outcomes, based on subjects' responses in the post-experimental questionnaire. We elicited responses to all possible outcomes after stage 1.[7]

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

[Figure 3 about here]

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 to their principals on average 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) estimates 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 0.6 lower in the DTG than in the TG.[9] Column (2) shows that if the outcome in the first stage is determined by a random mechanism instead of 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."

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

[9] The slopes of the trustworthiness curves in the DTG and the TG do not significantly differ from each other (the interaction between Trust and DTG is not shown).

Column (3) combines the two data sets and shows the difference of the differences. Trustworthiness is significantly less correlated with trust levels if a random mechanism determined the outcome in stage 1 than 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 (3) 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.

[Table 2 about here]

Taken together, the higher return ratio in the RDTG than the DTG, the lack of reciprocity and the absence of differential responses to omissions and to 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 trust, we find that only if the outcome in stage 1 is determined by a person rather than a random mechanism, reported trustworthiness is significantly lower in the DTG than in the TG. Agents "punish" principals harder for committing distrust than for omitting trust, but 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' (inaccurate) expectations.

Result 2 suggests that principals do not take agent framing effects into account 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. Table A.1 presents principals' expected trustworthiness (return ratio), specifically for the amounts of trust observed in the experiment and on average for all possible trust levels.[10]

When comparing expectations with the (experimental) return ratios for the amounts of trust principals chose in the experiment, we find that subjects are relatively more optimistic in the

[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.

DTG than in the TG. 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 to [*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 to [*y*/*x*]=0.93, p=0.17). The expected return ratios for the trust levels chosen in the experiment do not significantly differ from each other in the two games (p=0.60).[11]

Average expectations for all possible levels of trust, based on principals' strategy responses in the questionnaire, are quite similar in the DTG as in the TG. Principals expect an average return ratio of 0.96 in the DTG and 0.90 in the TG. 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).

Figure 4 shows expected return ratios for each possible level of trust and how expectations compare to reported return ratios. The DTG-slope of expected returns seems to be steeper than the TG-slope. A simple regression with fractions expected to be returned as the dependent variable, reported in Table A.2 in the Appendix, suggests that the slopes in the DTG and the TG significantly differ from each other. The correlation between trust levels and return ratio captures the intensity with which agents respond to a given level of trust. Principals expect that commissions lead agents to respond more strongly to a given trust level in the DTG than do omissions in the TG. Reported return ratios suggest a similar pattern but the difference in the reported trustworthiness slopes is not significant. Most importantly, principals fail to anticipate that commissions in the DTG decrease trustworthiness for all levels of trust compared to the TG.

[Figure 4 about here]

Table 3 shows that principals' (wrong) expectations are significantly related to their willingness to trust/not to distrust. Columns (1) and (3) shows that principals are more trusting in

[11] If return ratios and expectations thereof are calculated taking trust levels of zero ($x \geq 0$) into account, assuming E(*y*)=0 if *x*=0, subjects are still relatively more optimistic in the DTG than in the TG (DTG: E(*y*/*x*)=1.03 vs. [*y*/*x*]=0.34, p<0.01; TG: E(*y*/*x*)=0.63 vs. [*y*/*x*]=0.44, p=0.44). When we include trust levels of zero, the expected return ratios for the trust levels chosen in the experiment differ from each other in the two games (p<0.1).

the DTG than in the TG, for $x > 0$ and $x \geq 0$ respectively. When controlling for expected returns in Columns (2) and (4), however, this effect loses in importance and is no longer significant. Trust is only related to how optimistic principals are. While we cannot exclude the possibility that behavior affects expectations rather than the other way around with certainty, in Column (5), we use the game frame as an instrument for expectations.[12] Our results suggest that the game frames mainly affect trust by changing expectations.

[Table 3 about here]

While Table 3 does not suggest that framing exhibits a significant direct influence on behavior, expectations do not account for all of the variance in trust. We focus on the first stage of the TG and the DTG to examine loss aversion further. In the giving and taking dictator games, changing the default does not affect behavior. If 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. If 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.

Our results provide support for Conjecture 2b. Principals entrust the agent with more money in the DTG than in the TG because they expect higher returns in the DTG than in the TG for the experimentally observed trust levels. Our design does not allow us to conclusively rule out that loss aversion does not contribute to the higher trust levels in the DTG than the TG. 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 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

[12] Kosfeld et al. (Kosfeld, Heinrichs, Zak, Fischbacher and Fehr, 2005) exogenously increased trust levels and tested whether people adjusted their expectations accordingly. In their study trust levels were increased by randomly giving subjects oxytocin, a neuropeptide that plays an important role in social attachment and affiliation. Although the treatment group was more trusting than the control group, who received a placebo, expectations about agents' trustworthiness did not differ between the two groups.

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 without being trusted. They experience an increase in their earning of 130 percent compared to 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 the big corporations.

Why do principals not anticipate this? Our data suggests that they fail to walk in their agents' shoes. They are too optimistic about returns. Their inaccurate expectations lead them to trust too much in situations when they should not. 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.

Our principals are not able to accurately take their agent's perspective. Our findings are in line with a relatively large literature on biased perspective taking in psychology and most closely related to what van Boven et al. (2000) refer to as "egocentric empathy gaps," where people

overestimate the similarity between their own and others' valuations of a good, neglecting the endowment effect. Principals do not perceive distrust as a commission, which agents will punish more heavily than an omission. If anything, they are slightly more optimistic about returns when deciding not to distrust than when deciding to trust. For principals, not distrusting is easier than trusting as the former corresponds to the status quo while the latter requires departing from the default. They seem to expect that, analogically, it is somewhat easier for agents to reward lack of distrust than 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 obviously is 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 might generally 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 (e.g., 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 fall-back to complete distrust, widely acknowledged in the popular press to have occurred after the recent scandals (e.g., 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 (1992). "Warm-Glow Versus Cold-Prickle: The Effects of Positive and Negative Framing in Cooperation in Experiments." *Quarterly Journal of Economics* 60(1), pp. 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), pp. 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, pp. 475-98.

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

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

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

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

Bohnet, Iris and Rachel Croson (2004). "Special Issue: Trust and Trustworthiness." *Journal of Economic Behavior & Organization* 55(4).

Bohnet, Iris and Steffen Huck (2004). "Repetition and Reputation: Implications for Trust and Trustworthiness When Institutions Change." *American Economic Review* 94(2), pp. 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), pp. 467-84.

Bolton, Gary and Axel Ockenfels (2000). "ERC - a Theory of Equity, Reciprocity and Competition." *American Economic Review* 90(1), pp. 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), pp. 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), pp. 1-25.

Callahan, David (2004). *The Cheating Culture: Why More Americans Are Doing Wrong to Get Ahead*. Orlando et al.: 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), pp. 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, pp. 352-70.

Cox, James C. (2004). "How to Identify Trust and Reciprocity." *Games and Economic Behavior* 46(2), pp. 260-81.

Economist, The (2005). "Attempts to Reform Accounts Are Creating Their Own Problems." July 28th, pp. 65.

Ensminger, J. (2000). "Experimental Economics in the Bush: Why Institutions Matter," C. Menard, *Institutions, Contracts, and Organizations: Perspectives from New Institutional Economics.* Cheltenham: Edward Elgar Publishing, 158-71.

Fehr, Ernst; Georg Kirchsteiger and Arno Riedl (1993). "Does Fairness Prevent Market Clearing? An Experimental Investigation." *Quarterly Journal of Economics* 108(May), pp. 437-60.

Fehr, Ernst and Klaus Schmidt (1999). "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114(3), pp. 817-68.

____ (2003). "Theories of Fairness and Reciprocity - Evidence and Economic Application," M. Dewatripont, L. P. Hansen and S. J. Turnovsky, *Advances in Economics and Econometrics - 8th World Congress, Econometric Society Monographs.* Cambridge: Cambridge University Press, 208-57.

Foreman, Judy (2001). "First Rule: Don't Hesistate 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, pp. 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 Knetsch and Richard Thaler (1986). "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." *American Economic Review* 76(September), pp. 728-41.

Kahneman, Daniel and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47(2), pp. 263-91.

____ (1982). "The Psychology of Preferences." *Scientific American* 246, pp. 160-73.

Kosfeld, Michael; Markus Heinrichs; Paul J. Zak; Urs Fischbacher and Ernst Fehr (2005). "Oxytocin Increases Trust in Humans." *Nature* 435(2 June), pp. 673-76.

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

Lorsch, Jay W.; Leslie Berlowitz and Andy Zelleke eds (2005). *Restoring Trust in American Business*. Cambridge, MA: 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), pp. 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), pp. 267-75.

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), pp. 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), pp. 1281-302.

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

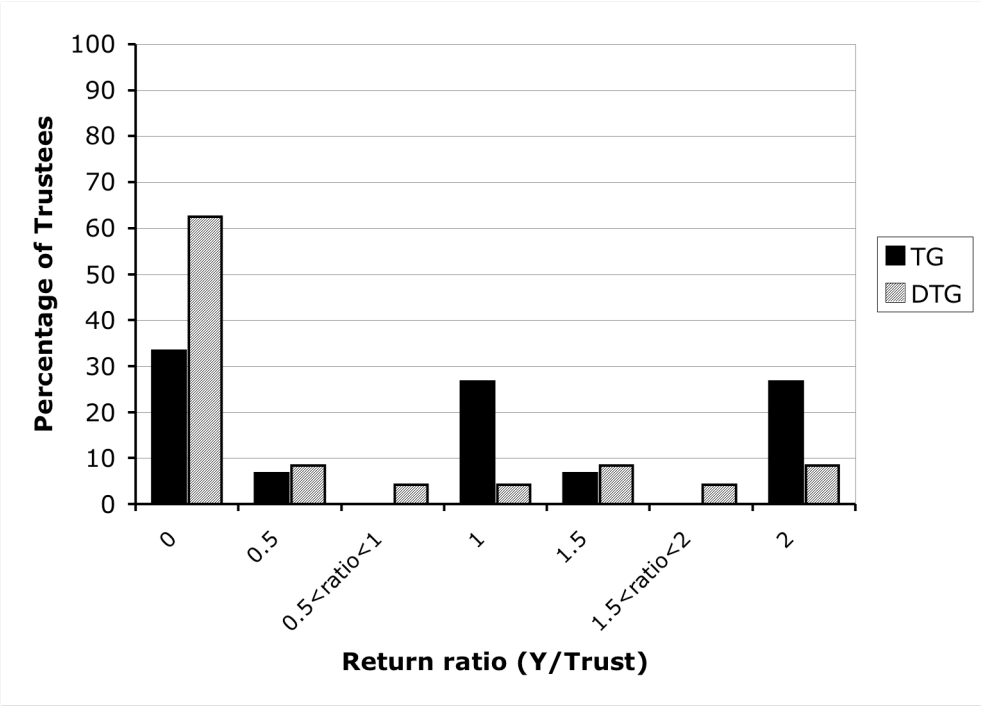
Simons, D.J. and C.F. Chabris (1999). "Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events." *Perception* 28, pp. 1059-74.

Thaler, Richard H. (1980). "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization* 1(March), pp. 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), pp. 66-76.

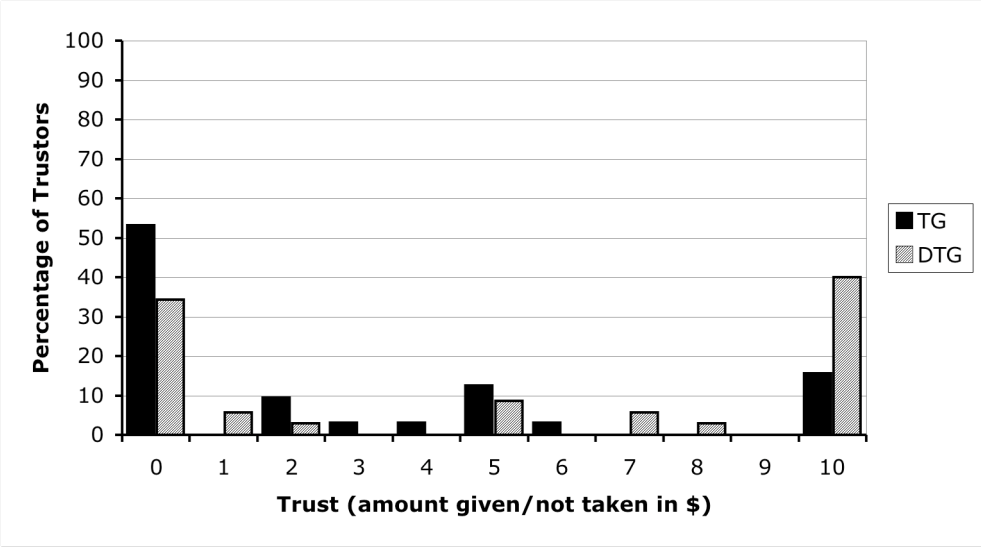
____ (2003). "Mispredicting the Endowment Effect: Underestimation of Owners' Selling Prices by Buyer's Agents." *Journal of Economic Behavior and Organization* 51(3), pp. 351-65.

Figure 1: Trustworthiness: Distribution of Agent's Return Ratios



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

Figure 2: Trust: Distribution of Amounts given by Principals



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

Figure 3: Trustworthiness: Reported Return Ratios

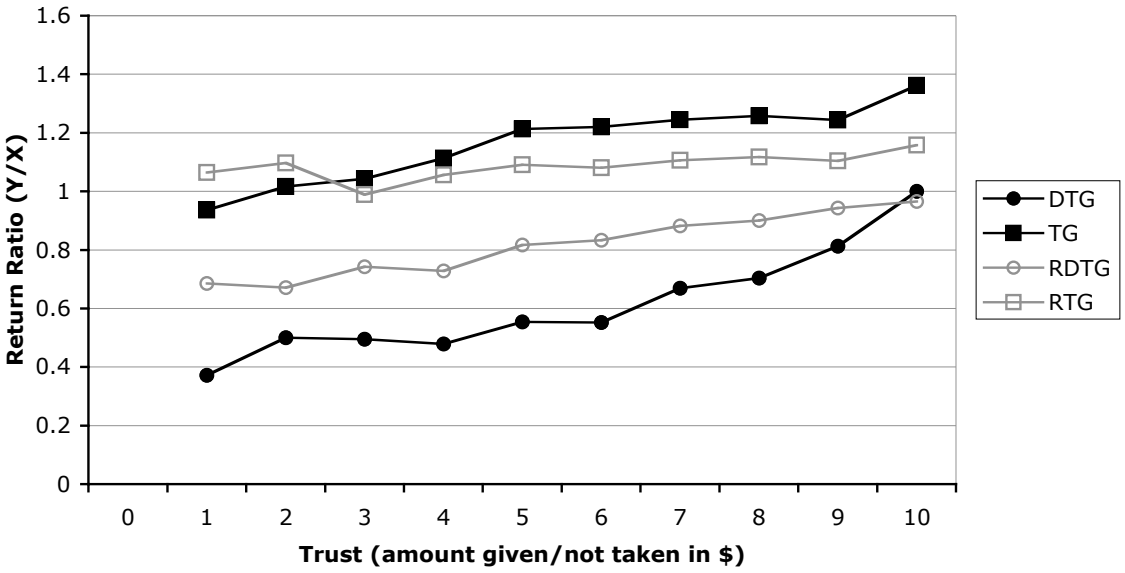


Figure 4: Expected and Reported Return Ratios

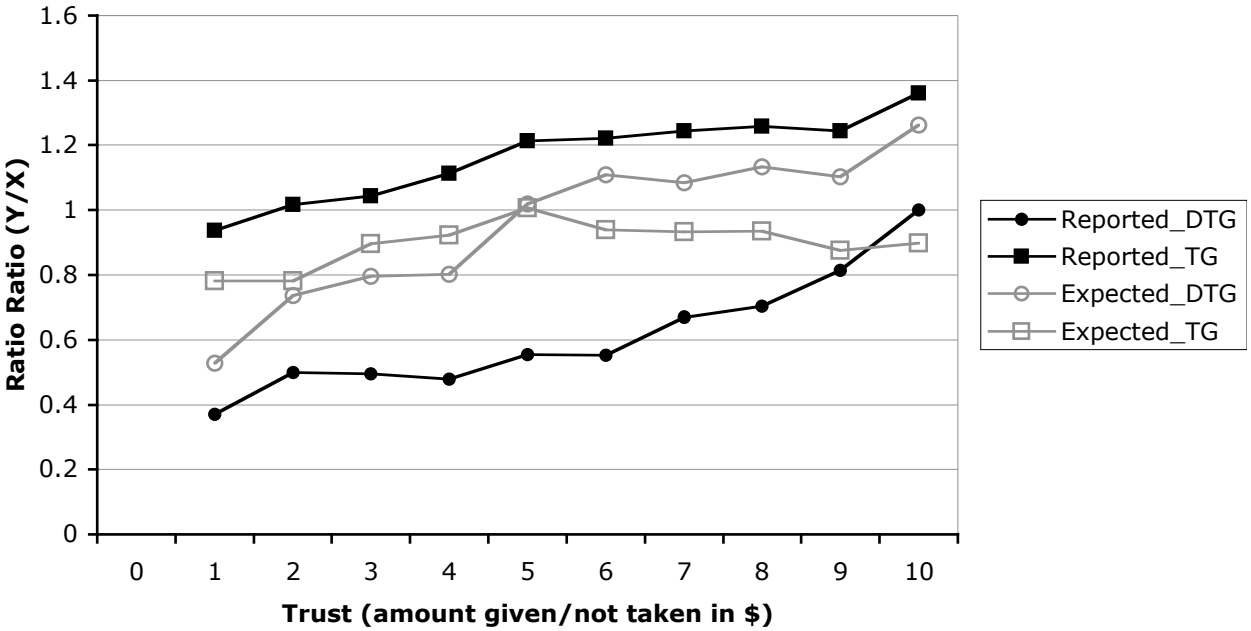


Table 1: Determinants of Trustworthiness

	(1)	(2)
Trust (amount given/not taken)	0.078(*)	0.096
	(0.039)	(0.062)
DTG (=1)	-0.573*	-0.373
	(0.259)	(0.602)
Trust*DTG		-0.030
		(0.081)
Constant	0.470	0.363
	(0.302)	(0.420)
Adj. R-squared	0.117	0.0947
# of observations	38	38

OLS-Regressions with standard errors in parentheses.
(*) 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) Random	(3) All
Given Trust Level	0.050**	0.023	0.050**
	(.013)	(0.017)	(0.013)
DTG (=1)	-0.551**	-0.269	-0.551**
	(0.186)	(0.214)	(0.186)
Random Allocation (=1)			0.070
			(0.150)
Random*Given Trust Level			-0.027*
			(0.011)
Random*DTG			0.282(*)
			(0.142)
Constant	0.890**	0.960**	0.890**
	(0.178)	(0.211)	(0.178)
R-squared	0.1203	0.0228	0.0679
# of observations	660	660	1320
# of individuals	66	66	66

Notes: 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)	(5)
DTG (=1)	1.936(*)	1.515	2.390*	0.823	
	(1.055)	(1.012)	(1.021)	(0.633)	
Expected Return Ratio		1.807*		3.837**	5.852**
		(0.779)		(0.358)	(1.828)
Constant	5.933**	3.484*	2.781**	0.343	-0.937
	(0.821)	(1.309)	(0.738)	(0.500)	(1.596)
Adj. R-squared	0.0602	0.162	0.0636	0.660	0.499
# of observations	38	38	67	67	67

Notes: Models (1)-(4) show OLS-regressions with standard errors in parentheses. Model (5) shows a 2SLS regression with standard errors. First-stage regression: Expected Return Ratio = 0.408 (s.e. = 0.213) DTG + 0.635 (0.154) constant. Adj. R-squared: 0.039. Models (1) and (2) for $x > 0$; models (3)-(5) for $x \geq 0$.

(*) 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
Return Ratio by Agent, y/x, for $x > 0$ (Trustworthiness)	0.933 (0.821) [15]	0.51 (0.746) [23]	|z| = 1.747 p = 0.081
Return Ratio by Agent, y/x, for $x \geq 0$ (Trustworthiness)	0.438 (0.727) [32]	0.336 (0.649) [35]	|z| = 0.600 p = 0.548
Amount returned by Agent (y)	3.063 (5.842) [32]	3.057 (6.058) [35]	|z| = 0.520 p = 0.603
Trust (amount given/not taken) by Principal(x)	2.781 (3.705) [32]	5.171 (4.560) [35]	|z| = 2.130 p = 0.033
Average earnings by agents	15.281 (7.646) [32]	23.314 (12.153) [35]	|z| = 2.747 p = 0.006
Average earnings by principals	10.281 (3.846) [32]	7.886 (5.556) [35]	|z| = 2.251 p = 0.0244
Efficiency (Average earnings by pairs)	25.563 (7.409) [32]	30.343 (9.120) [35]	|z| = 2.130 p = 0.033
Questionnaire			
Reported Return Ratio (with human principal)	1.165 (0.151) [31]	0.614 (0.112) [35]	|z| = 2.982 p = 0.0096
Reported Return Ratio (with random mechanism)	1.086 (0.171) [31]	0.817 (0.130) [35]	|z| = 0.950 p = 0.342
Average Expected Return Ratio by Principal	0.896 (0.147) [32]	0.957 (0.119) [34]	|z| = 0.485 p = 0.628
Expected Return Ratio for Chosen Amount of Trust; for $x > 0$	1.356 (0.796) [15]	1.589 (0.521) [23]	|z| = 0.519 p = 0.604
Expected Return Ratio for Chosen Amount of Trust; for $x \geq 0$	0.635 (0.154) [32]	1.030 (0.151) [34]	|z| = 1.718 p = 0.086
Amount given/not taken in Dictator Game	0.438 (0.220) [32]	0.471 (0.308) [34]	|z| = 0.136 p = 0.892

Notes: For $x = 0$, return ratios are assumed to be 0. 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:
$ _____	→	$ _____	$ _____