Trust, Reciprocity, And Interpersonal History: Fool Me Once, Shame on You, Fool Me Twice, Shame on Me

By

John Dickhaut, University of Minnesota
Kevin McCabe, George Mason University
Radhika Lunawat, University of Minnesota
John Hubbard, Macalester College

September 2008

Correspondence to:

John Dickhaut, Department of Accounting
University of Minnesota, Carlson School of Management
321-19th Avenue South, Room 3-113, Minneapolis, MN 55455-0413
JDickhaut@cson.umn.edu

The authors acknowledge support from National Science Foundation Grant #SBR-9300287, National Science Foundation Grant #SES-9210052, and the Accounting Research Center at the University of Minnesota.

Abstract

We design an experiment to study the effect of reputation building on trust and reciprocity in a two period investment game. In the investment game a first mover decides how much money (up to $10) to invest. It is common knowledge that the amount invested will triple by the time it reaches the second mover. The second mover must then decide how much of the tripled money to keep and how much to return. The money returned does not triple a second time. If the investment game is played once, noncooperative game theory predicts that the second mover has a dominant strategy (under the assumption of self-interest) to return zero. If self-interest is common knowledge, then first movers should invest zero. When the investment game is played once, with double blind anonymity conditions, Berg, Dickhaut, and McCabe (1995) find high levels of both investment and amounts returned. These results are interpreted as behavioral evidence for the existence of trust and positive reciprocity.

In this paper we study the value of reputations, as trusting or trustworthy types, in promoting first mover investment in sequential exchange. To study reputation building the investment game is played twice. In both periods subjects keep the same partners and roles. Data for twenty-three pairs of subjects is reported. In period one, amounts sent and amounts returned increases significantly over the already high baseline levels found in one-shot play. In period two, the amounts of reciprocity falls to the baseline level found in one-shot play. Maximum likelihood estimates of types shows that the data is most consistent with the predictions of a Bayesian Nash equilibrium where the existence of 'trusting' types leads subjects to build reputations as 'trustworthy' types.

1. Introduction

In this paper we study the value of reputations in promoting first mover investment in sequential exchange. Examples of sequential exchange are numerous including, capital markets, where investors invest capital before cash returns are realized, and labor markets, where employees invest time and effort before wages are paid. We are interested in those cases where the division of the monetary gains from exchange is determined after investments have been made, by someone else, e.g., stewards or employers, who presumably prefer more gains to less. Given the capriciousness of future returns, it is difficult to understand why first movers are willing to participate at all.

In his lectures on the limits of organizations, Anow (1974) notes that in the face of transactions costs trust is ubiquitous to almost every economic transaction. This question raises important questions about economic behavior. Is trust a behavioral primitive which needs to be incorporated into economic theory? What factors increase or decrease the use of trust? In previous work Berg, Dickhaut, and McCabe (1995) (herein after BDM) present experimental evidence for the existence of trust. But, the existence of trust allows for the possibility that subjects will respond by building reputations as trusting (or trustworthy) types. In our experimental design, reputations increase efficiency by promoting trust relations.

In Coleman's (1990) terminology a trust relation is defined in terms of two actions: first, a trustor 'places a trust' by giving a trustee the right to make a decision; and, second, the trustee makes a decision which affects both trustor and trustee. If the trustee's decision makes both players better off, relative to the initial state, then the trustee is said to 'return the trust.' To interpret the resulting transaction as one that has used trust, the following conditions are sufficient:

(1) Placing trust in the trustee puts the trustor at risk;

(2) Relative to the set of possible actions, the trustee's decision benefits the trustor at a cost to the trustee; and

(3) Both trustor and trustee are made better off from the transaction compared to the outcome which would have occurred if the trustor had not entrusted the trustee.

Coleman's definition makes it possible to study trust relations in terms of trusting and trustworthy types without explaining how these types emerge.

The investment game, studied by BDM*, provides an elementary example of a trust relation between two players. In the investment game, player one places a trust by deciding how much of ten dollars to send to player two. It is common information that the amount sent will triple before it reaches player two. Player two then returns the trust deciding how much of the tripled money to keep and how much to return.

The unique Nash equilibrium for the investment game is for player one to send nothing. In particular, player two has a dominant strategy (under the usual assumption of self-interest) to return zero. Given player two's dominant strategy, player one's best response is to invest zero.

---

*The investment game is similar to the trust game presented in Kreps (1990), the centipede game in Rosenthal(1982), and the peasant-dictator game studied by Van Huyck, Battalio, and Walters (1993). While the trust game and the centipede game have two choices at each stage, the investment game has a larger choice space allowing for different degrees of trust and reciprocity. All of these games have the same noncooperative prediction that play should end immediately even though strict Pareto improvements to payoffs can be found in later stages. McKelvey and Palfiey (1992) study repeated, one-shot, plays of four and six stage centipede games where subjects are paired with a different counterpart at the end of each game. Subjects in centipede games to cooperate thus reaching later stages. McKelvey and Palfhy that a game of incomplete information based on a reputation for altruism i.e., always willing to continue to later stages, explains their data.

The unique Nash equilibrium for the investment game is for player one to send nothing. In particular, player two has a dominant strategy (under the usual assumption of self-interest) to return zero. Given player two's dominant strategy, player one's best response is to invest zero.

When the investment game was played once, under double blind anonymity conditions, with thirty-two inexperienced subjects BDM find that 30 of 32 subjects sent a positive amount averaging $5.16. A third of the time player 2 reciprocated by returning more than was sent. Since the double blind controls eliminate the possibility of reputations, contractual precommitments, and the threat of punishment, i.e., negative reciprocity, these results provide strong behavioral evidence for the existence of trust. BDM refer to this treatment as No History since subjects had no history with respect to each other in the play of the investment game

Are the amounts sent in the No History treatment motivated by trust? Perhaps subjects send money because they have not thought through the dominant strategy of player 2's or perhaps they are curious and are willing to pay to satisfy their curiosity. BDM run a second treatment, Social History, which deals with these concerns. In this treatment a second group of subjects is given the history of both the amounts sent and amounts returned by the first group of subjects. BDM find that social history strengthens the relationship between trust and reciprocity. Lack of understanding, or curiosity, cannot explain this behavior, since player one's no longer have to send money to find out what might happen. Instead BDM conjecture that a social history acts to reinforce shared social norms thus making reciprocity more likely.

We can explain trust in the investment game by introducing two types of players. Player one is 'trusting' if he or she begins by investing, but stops as soon as player two fails to reciprocate. This kind of behavior is similar to the cooperative tit-for-tat strategy in prisoners' dilemma games. Player two is 'trustworthy' if he or she reciprocates by returning more than was sent. Note that trusting behavior can be explained as a calculative response to a belief in trustworthy types. The more difficult problem is to explain the existence of trustworthy behavior.

Frank (1988) introduces a strategic role for the emotions as a mechanism (either biological or societal) for allowing reciprocity contrary to self interest. According to Frank's commitment model reciprocity occurs (even in single play) because trustworthy types have refined their emotional responses to favor this behavior. The commitment model requires that such refinement is developed through a costly process of learned self control and therefore is difficult to mimic by persons who yield to their psychological impulses for immediate return. Thus in a one-shot game trustworthy types exist because succumbing to self interest now makes it more difficult to resist temptation later.

Given that observed behavior can be explained by trusting and trustworthy types it is possible to analyze this behavior using incomplete information games. In particular, do people try to build reputations as 'trustworthy' or 'trusting' types? The design of this experiment is similar to that used by BDM, except subjects play the investment game twice with the same counterpart. Standard noncooperative game theory with complete information continues to predict that interpersonal history should not have an impact on investment decisions or reciprocity. Alternatively, if we allow subjects to have incomplete information about their counterpart's type, then non-cooperative game theory predicts an increase in trust and reciprocity as people build reputations and expectations on each others types based on their expectations and experience. See Kreps (1990)

Data for twenty-three pairs of subjects supports the reputation based prediction. In period one, trust and reciprocity increase significantly over one-shot play observed in BDM. In period two, reciprocity falls significantly to levels similar to those found in the one-shot play. Consistent with the predictions of a Bayesian Nash equilibrium, room B subjects send back significantly more in period one compared to period two. A maximum likelihood estimate of the proportion of 'trusting' and 'trustworthy' types finds a significantly higher proportion of trusting types thus supporting the interpretation that reputations are built on the existence of 'trusting' types.

2. Experimental Design

The two period investment game is played as follows. In period one, subject a, in room A, is given \$10 as half of a show up fee. He or she then decides how much of this money to invest with an anonymous counterpart in room B. We denote this amount by m_1, which can take any of the discrete values {\$0, \$1, . . . , \$10}. The amount invested then triples, resulting in a total return of $3m_1$. A counterpart in room B, called subject b, who has also been given \$10 as a show up fee, receives the tripled amount $3m_1$, and then decides how much of the $3m_1$, to keep and how much to send back to subject a. The amount kept is denoted k_1, which can take any of the discrete values $\{0, 1, \ldots, 3m_1\}$. Subject a receives the amount, $[3m_1 - k_1]$.

In period two, subject a is given another \$10, the second half of a show up fee, and decides how much of it to invest with the same anonymous counterpart. We denote this amount by m_2, which can take any of the values {SO, \$1, . . . , \$10). The amount sent then triples, resulting in a total return of $3m_2$. Subject b, who is also given an additional \$10 as a show up fee, receives $3m_2$ and then decides how much to keep and how much to send back to subject a. The amount kept is denoted k_2, which can take any of the values $\{O, l, \ldots, 3m_2)$. Subject a then receives the amount, $[3m_2 - k_2]$.

Since the two period investment game consists of four distinct decision dates it is possible that this additional complexity may lead to behavior due only to inexperience. To control for this issue we use two distinct groups of subjects, an experienced group which previously played in the one shot investment game, reported in BDM, and an inexperienced group with no previous experience in the investment game. The variable x will distinguish experience levels, where x = 0 indicates inexperienced subject data, and x = 1 indicates experienced subject data.

We refer to the BDM, No History, baseline results by the subscript N. For the two period investment game, we refer to period one results by the subscript I = 1, and period two results by the subscript I = 2. For each case we describe the cumulative density function for the subject

population as follows. The c.d.f on amounts sent in period I, for experience level x, is denoted $S_{i,x}(.)$. Similarly, the c.d.f. on amounts returned, is denoted $R_{i,x}(.)$. For the No History baseline only inexperienced subjects were used in a one-shot game. We denote these cumulative by $S_n(.)$ and $R_n(.)$.

Since trust and reciprocity are observed in the BDM (No History) baseline it may be the case that subjects will be unaffected by their two period play with the same partner. lhis observation leads us to use the following Null Hypotheses that the level of trust and reciprocity in both the first and second periods will be the same, and that these levels win be the same as the baseline levels found in BDM.

H_0: $S_{1,x} = S_{2,x}$ amd $R_{1,x} = R_{2,x}$.

H_0': $S_N = Si_{,x}$ and $R_N = R_{i,x}$, for i=1,2.

Consistent with the idea that the investment game is played with incomplete information on a counterpart's type we introduce a trustworthy type. Subject b is trustworthy if he exhibits positive reciprocity, i.e., $k_t = 3/2\ m_t$ for t=1,2. This type can be distinguished from a self interested expected utility maximizer who may act as a trustworthy type in period one but will keep all the money in period two. Let pt be subject a's belief on the probability of subject b being trustworthy. If $p_t(3/2\ m_t) > m_t$, and subjects are risk neutral, then it is clearly in subject a's interest to send money. Thus if $p_t > 2/3$, subject a should send m_t = \$10.

Let p_1 be the common knowledge priors of participants in the investment game. Note, these priors will be updated at the end of period one based on subjects' analyses of their counterparts behavior. One model of this updating rule is given by the definition of a Bayesian Nash equilibrium. If $p_1 > 2/3$, then the following Bayesian Nash equilibrium predicts that untrustworthy types will build a reputation as trustworthy in period one.

I. Trustworthy (room B) types: $k_t = 3/2\ m_t$, for t = 1,2.

Untrustworthy (room B) types: $k_1, = 3/2\ m_1$, $k_2 = 0$.

Room A players: $m_1 = 10$, if $k_1 > 3/2\ m_1$ then $m_2 = 10$, else $m_2 = 0$.

To see that I is a Bayesian Nash equilibrium we see that an untrustworthy subject b who follows this strategy is maximizing expected payoff. First note that $p_1 > 2/3$ implies that both types of subjects a's will send $m_t = 10$ as long as $p_t = p_1$. However, if subject b fails to reciprocate then both type a's update $p_2 = 0$ and send nothing in period two. Following the equilibrium strategy results in \$15 + \$30 = \$45 for untrustworthy subject b's. Since it is always optimal to defect in period two, the other possibility is to defect in period one, i.e., $k_1 = k_2 = 0$. But, k1=0 informs subject a that subject b is untrustworthy causing subject a to update $p_2 = 0$. In this case subject a's best response in period 2 is to keep the \$10 and send nothing. Subject b's total payoff is \$30 + \$0 < \$45.

A strong prediction for the reputation model is for subject a's to invest \$10 in period one since they will receive \$15 for sure. However, there is evidence in BDM that subjects may be more heterogeneous in both their beliefs and level at which they reciprocate. For example, if subjects are unsure of this equilibrium there may be a tendency to reduce m1. While the reputations model makes strong assumptions on types, we introduce it to illustrate how reputations are predicted to form in equilibrium. If, however, types are more heterogeneous, this creates a large number of information sets in the formal analysis making it difficult to believe that subjects can mutually achieve a sequential equilibrium in one play of the two period game. For this reason we state the null hypothesis in the following weaker forms.

H1: $S_{1,x}=S_{2,x}$ and $R_{1,x}>R_{2,x}$;

H1': $S_{2,x}=S_n$ and $R_{2,x}=R_n$

We can also define a type for room A subjects as follows. Subject a is trusting if subject a sends \$10 in period one, and if the period one decision is reciprocated, sends \$10 in period one, and if the period one decision is reciprocated, sends \$10 in period two. If however, the period one decision is not reciprocated, trusting types will send 0 in period two. This type is consistent with the tit-for-tat and other strategies for cooperation in repeated games which give cooperation the benefit of the doubt. Let $(p_1\ q_1)$ be the common knowledge priors of participants in the investment game. Note, these priors will be updated at the end of period one based on subjects' analysis of their counterparts behavior. One model of this updating rule is given by the definition of a Bayesian Nash equilibrium. Depending on the values of $(p_1\ q_1)$ satisfies $p_1>2/3$, then the type I equilibria predict that untrustworthy types will build a reputation as trustworthy in period one.

If (p_1, q_1) satisfies $q_1>1/2$, then the following Bayesian Nash equilibrium predicts that untrusting subject a's will also build reputations as trusting.

II. Trustworthy (room B) types: kt=3/2 mt , for t=1,2.

Untrustworthy (room B) types k1=3/21, k2=0.

Trusting (room A) types: m1=10, if k1>10, then m2=10, else m2=0

Untrusting (room A) types: m1=10 and m2=0.

To see that II is a Bayesian Nash equilibrium, note that subject b will update q_2 as follows: if $m_1>0$ then $q_2=0$. Subject a's will update p_2 as before. If $q_2>1/2$ and $m_2\geq m_1$, then $q_2(3/2m_1+3m_2)+(1-q_2)3/2m_1>3m_1$. An untrusting type should send \$10 in period one since this results in \$15 and keep their money in period two. Their best alternative is to send nothing in

both periods, but \$25 > \$20. The strong predictions of type II equilibria again predict \$10 send and reciprocity in period one, but cheating occurs by the untrusting type a's, and untrustworthy type b's, in period two. The weaker form of the reputation prediction allows us to restate our first alternative hypothesis as follows:

H1: $S_{1,x} \geq S_{2,x}$ and $R_{1,x} > R_{2,x}$

We will u se the Kruskal-Wallis test to examine the between subject design hypothesis that levels of trust and reciprocity in each period of the two period game are no different from the no history baseline, i.e., comparing H_0' and H_1'. We will then use an ordered Wilcoxon test to examine the within subject design hypotheses that levels of trust and reciprocity in the two period investment game, controlling for experience, are the same across periods, i.e., comparing H_0 and H_1. If both these tests support their respective null hypotheses, then we must conclude that we find no support for the reputation model. If both these tests reject their respective null hypotheses, then we will conclude that reputation building does exist.

3. Procedures

Experience subjects were recruited by phone from undergraduate student population at the University of Minnesota. Except for prospective monitors, every subject chosen had participated in a previous trust experiment (either the no history or the social history treatment). Inexperience subjects were recruited from summer school classes at the University of Minnesota. Except for the prospective monitors, none of these subjects had participated in a previous trust experiment. Separating subjects by their previous experiences in similar experiments allows for the measurement of differences between the groups.

Subjects were told to report directly to either room A room B, and were moved to a different room if needed. Once at least four subjects were in each room, monitors were randomly

chosen from the subjects in the room. One subject was bumped due to an odd number of show ups, and was given $5 as a show up fee. Experiments lasted from sixty to ninety minutes.

A. The Double Blind Procedure

In the investment game subject a places a trust by sending money. Suppose instead that experimenter gives subject b $10 with the understanding that subject b must then decide how much to keep and how much to give (not return) to subject a. This game, called the dictator game, raises the following question. How will subject b behave when subject a has not placed a trust? The dictator game is studied in Forsythe, Horwitz, Savin and Sefton (1994). They find that subjects send significant amounts of money to their counterparts, i.e. only 18% send $0 while 32% send $4 or more to their counterpart. Since this behavior is not due to trust it can confound the other-regarding behavior of room B subjects.

Hoffman, McCabe, Shachat, and Smith [1994] find that a double blind procedure (Double Blind 2) intended to guarantee almost complete social isolation of the individual's decision (no one including the experimenter or any subsequent observer of the data could possibly know a subjects decision except in the unlikely even that every subject makes the same decision) significantly reduces this confound, i.e., 58% now send $0 while only 12% send $4 or more. In subsequent work Hoffman, McCabe, and Smith [1995] systematically confirm the effect of social isolation, which they call Social Distance, on reducing the confound on other-regarding behavior. This work motivates using Double Blind procedures for the study of trust.

In implementing the two stage investment game we extended the double-blind procedures used by BDM. Our choice was motivated by the desire to keep all decisions anonymous. Not even the experimenter is able to map individuals decisions made to the identity of the decision maker, yet it is still possible to gather partners data. Details of this procedure can be read in the instructions.

3. Experimental Data

Twenty-three pairs of subjects were run: 12 of these pairs were subjects who had taken part in previous trust experiments, and 11 of these pairs had no such experience. In Appendix B we list our data from these experiments.

For the subjects with previous experience in a trust experiment, the average amount sent from room A in the first stage was \$6.58, producing an average return of \$8.92, compared with an average amount of \$4.75 send in the second stage, which resulted in an average return of \$5.00. For subjects with no experience in a trust experiment, the average amount sent from room A in the first stage \$7.09, resulted in an average return of \$12.36, while \$7.18 was sent on average in the second stage, producing an average return of \$5.18.

A. Period One Results

In period one, subjects chose amounts to send, m_1, and keep, k_1, knowing that another stage of the same process would follow. Figure 1 graphs the results from this treatment. The data is presented in descending order using the amounts sent from room A (m_1,show in open circles) as the primary key. Ten of the 23 subjects in room A send the maximum amount allowed, \$10. Gray bars indicate the total amount received by a counterpart in room B. Thus when \$10 was sent, \$30 was received by a counterpart in room B.

The date is further sorted in descending order using amounts returned ($(3m_1-k_1)$, shown as black circles) as the secondary sort key. When amount returned is greater than amount sent, the net return is positive. For example, the highest payback was \$30 on an amount sent of \$10, resulting in a net return of \$20. The lowest payback on \$10 sent was \$5, resulting in a net return of -\$5. Asterisks mark those pairs without experience in a previous trust experiment.

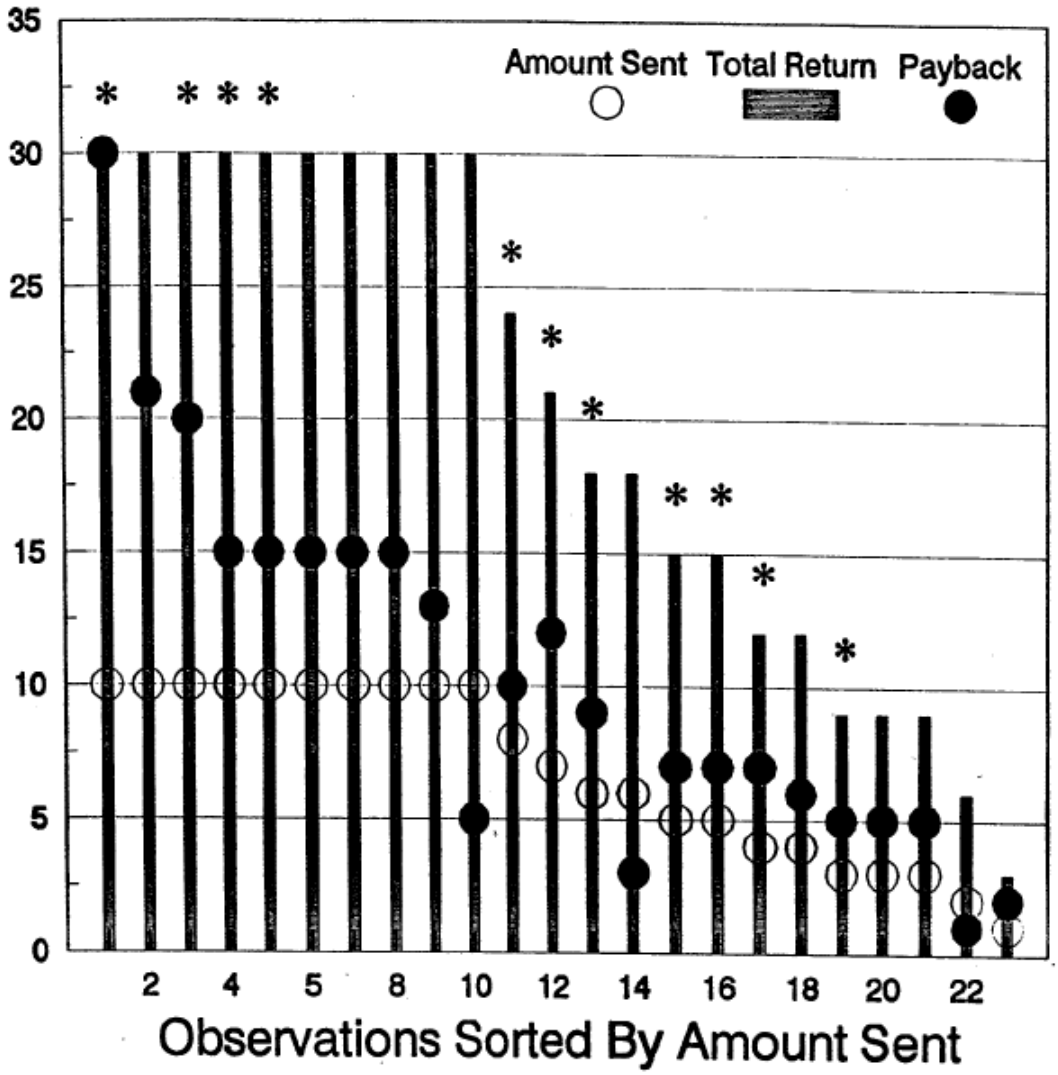
We make the following observations. For room A decisions; (1) None of the 23 subjects sent zero, and (2) Amounts sent exhibit high variability with a mode of 10 subjects sending $10. For room B decisions; (3) None of the 23 room B subjects returned zero. However, three subjects returned amounts resulting in negative returns. (4) Twenty subjects in room B returned more than their counterpart sent, resulting in positive net returns. Overall, an average amount sent of $6.83 resulting in an average amount returned of $10.57, and average net return of $3.74.

B. Period Two Results

In period two, subjects have the experience of having just participated in the first period with the same counterpart. This allows us to measure the effects of interpersonal history on amounts sent. Figure 2 graphs the results from these sessions. The date is sorted and presented similarly to the date in Figure 1. The amounts sent from room A (m_2) are shown in open circles. Gray bars indicate the total amount received by the counterpart in room B. Amounts retuned ($3m_2$-k_2) are shown as black circles.

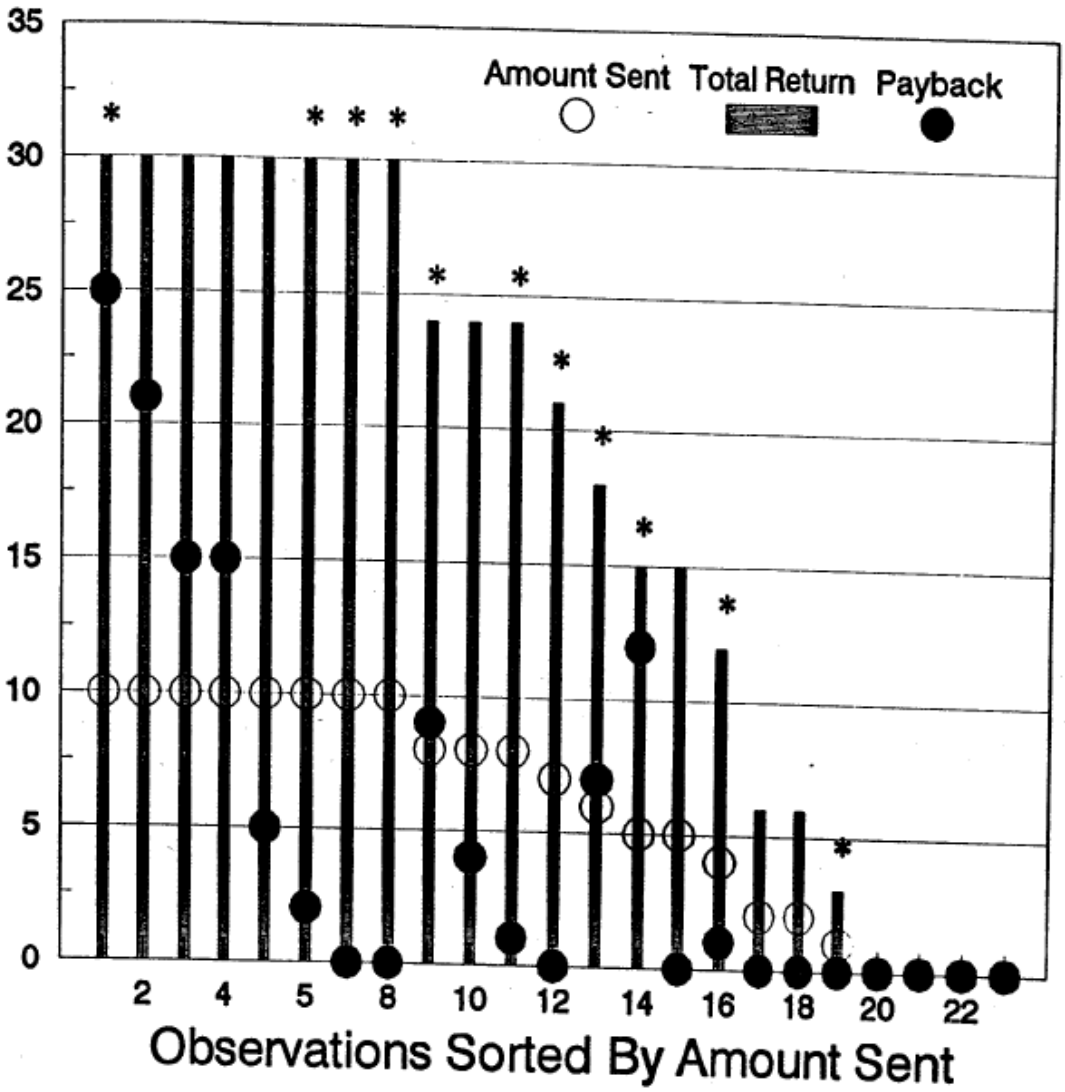
We make the following observations. In room A; (1) we continue to observe a large degree of variability with four of the subjects sending zero dollars, and a mode of eight subjects sending $10. In room B; (2) eleven of the twenty-three subjects returned zero dollars. Although four of these subjects had no payoffs to return, the seven others kept the total return, consistent with the loss of reciprocity hypothesis, A_1. (3) Twelve subjects returned positive amounts, but only seven of these resulted in positive net returns to the people in room A. The latter behaviors provide some support for hypothesis A_3. In the second stage, the average investment was $5.91 and the average payback was $5.09, resulting in an average net return of-$0.83.

Figure 1
Trust With Inter-Personal Histories (Period 1)
Amount Sent, Total Return, and Payback



Maximum Investment is $10
Maximum Return is $30
Payback Less Than Investment Implies Loss
* = Inexperienced Subjects

Figure 2
Trust With Inter-Personal Histories (Period 2)
Amount Sent, Total Return, and Payback



Maximum Investment Is $10
Maximum Return Is $30
Payback Less Than Investment Implies Loss
* = Inexperienced Subjects

C. Results Across Periods

Figure 3 shows partners decisions for both periods. The data is sorted and presented similarly to the data in Figure 1. The amounts sent from room A, $(m_1\text{-}m_2)$, are shown in open circles. Gray bars indicate the total amount received by the counterpart in room B. Amounts returned $[(3m_1\text{-}k_1)+(3m_2\text{-}k_2)]$ are shown as black circles. Six room A subjects invested \$10 both periods. None of these players makes a loss. As the total amount invested decreases we observe more overall negative returns. However, since most partners experienced net gains in period one, losses are less than those observed in the one-shot (no history) treatement found in BDM.

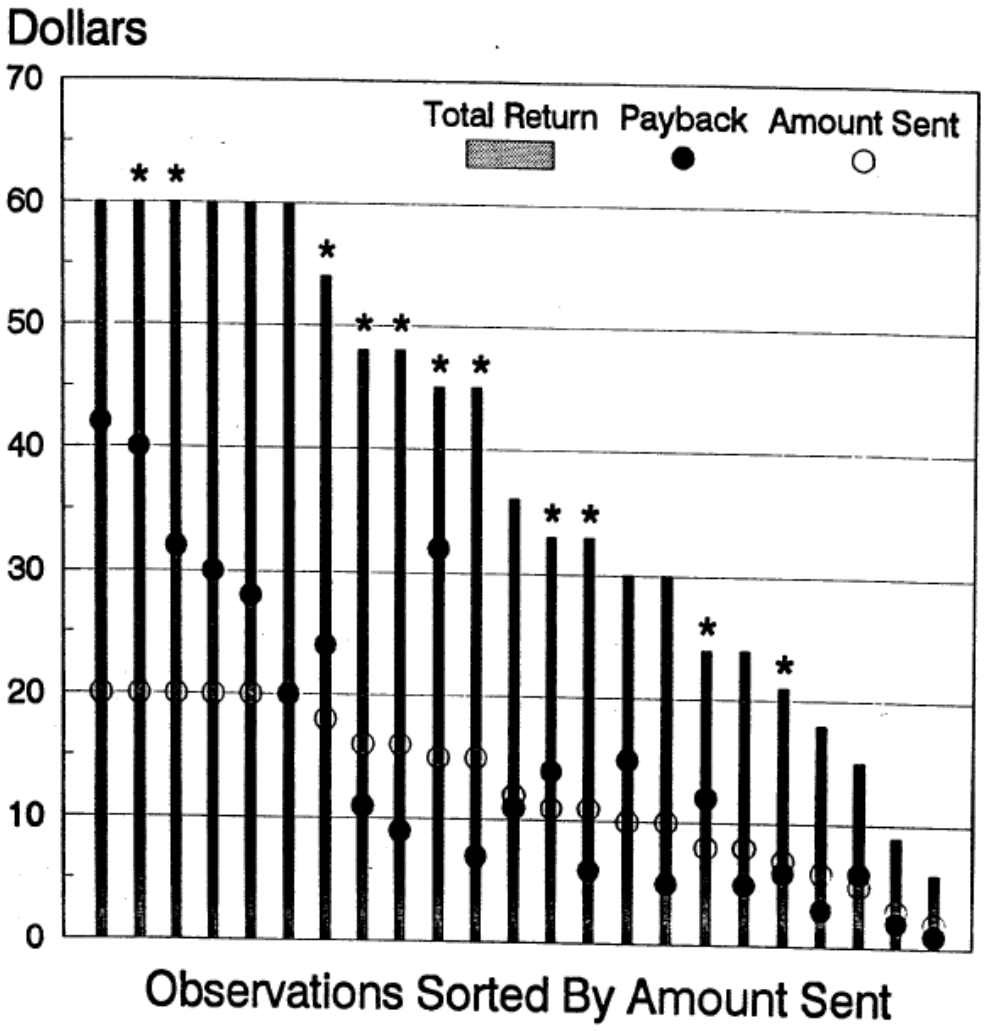
Consistent with the reputation hypotheses, all three persons from room A, who received a negative net return in period one, sent zero to their respective counterparts in period two. Of the twenty people who received positive returns in the first period, nineteen invested positive amounts in the second period. However, in comparison to the twenty out of twenty-three room A subjects who received positive net returns in the first stave, only seven of the nineteen room A subjects received positive net returns in the second stage. This marks a sharp decrease in reciprocity in the second stage.

D. Test of Hypotheses

Figure 4 shows box plots of subjects decisions broken down by period and experience level. Amounts sent are measured as a percentage $m_t/10$, while amount returned are measured as a percentage return $(3m_t\text{-}k_t)/3m_t$. Since some subjects in room B received zero in period two we do not know what their return decision might have been. These data points are treated as missing observations in subsequent analysis. The box plots show a large degree of variability in all cases except percent return in period 1 which is distributed narrowly around $3/2m_1$ or 50%. This is

Figure 3

Trust With Inter-Personal Histories
Totals For Both Periods



Maximum Investment is $10
Maximum Return is $30
Payback Less Than Amount Sent Implies Loss
* = Inexperienced Subjects

Figure 4

Box Plots Of Amounts Sent And Percentage Returns by Experienced Subjects vs. Inexperienced Subjects

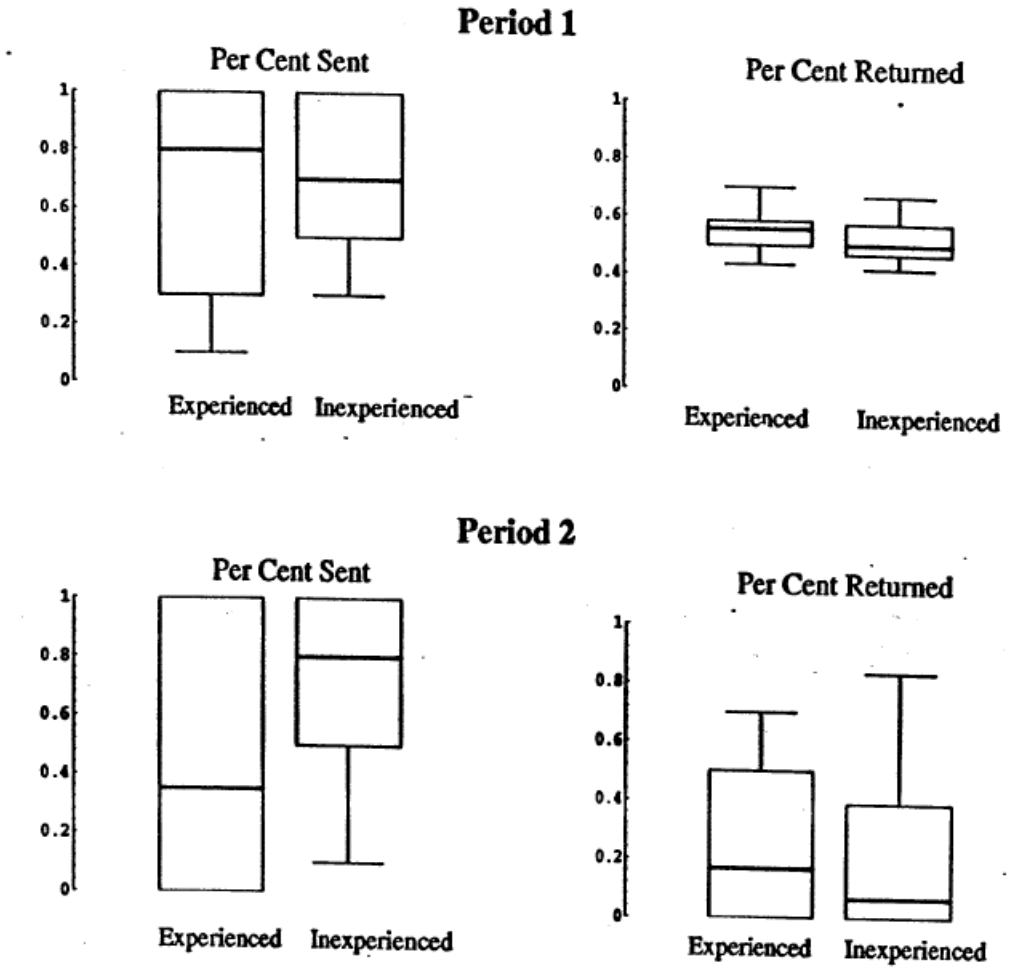


Table 1

Statistical Comparisons Of Treatment Effects
(KW = Kruskal-Wallis Statistic, Z = Normalized Wilcozon z)
(P = Probability of Accepting the Null Hypothesis of No Effect)

A. Comparisons of Amounts Sent

	Exp. Period 1 (N = 12)	Inexp. Period 1 (N = 11)	Exp. Period 2 (N = 12)	Inexp. Period 2 (n = 11)
No History (N = 32)	KW = 1.305 (P = 0.25)	KW = 3.497 (P = 0.06)	KW = 0.273 (P = 0.60)	KW = 4.262 (P = 0.04)
Exp. Period 1 (N = 12)		KW = 0.097 (P = 0.76)	Z = 17.50 (P = 0.83)	
Inexp. Period 1 (N = 11)				Z = 26.50 (P = 0.72)
Exp. Period 2 (N=12)				KW = 1.489 (P = 0.22)

B. Comparisons of Percentage Returns

	Exp. Period 1 (N = 12)	Inexp. Period 1 (N = 11)	Exp. Period 2 (N = 8)	Inexp. Period 2 (N = 11)
No History (N = 32)	KW = 6.24 (P = 0.012)**	KW = 6.67 (P = 0.01)**	KW = 0.163 (P = 0.69)	KW = 0.758 (P = 0.38)
Exp. Period 1 (N = 12)		KW = 0.759 (P = 0.38)	Z = 3.00 (P = 0.02)	
Inexp. Period 1 (N = 11)				Z = 7.00 (P = 0.01)
Exp. Period 2 (N = 8)				KW = 0.062 (P = 0.80)

**These treatment comparisons are significant at 5%.

consistent room A subjects mimicking (or pooling with) 'trustworthy' types.

In table 1 we report our statistical tests between treatments. For our samples we find no significant experience effects at the .05% or .1% levels for amounts sent or percentage returns in period 1 or period 2. However, we do observe weak statistical support for a difference in amounts sent by inexperienced subjects and amounts sent in our no history baseline.

We do get strong statistical support for our reputation hypothesis. While we cannot reject the null hypothesis that amounts sent are the same in both periods, we can reject the null hypothesis that percent returned is the same in both periods in favor of the alternative hypothesis that percent return is higher in period one. Further consistency with the reputation hypothesis is that percent returns in period 2 are not different from percent returns in no history. Our conclusions are strengthened when we pool inexperienced and experienced subjects as shown in Table 2.

5. Estimation Of Types In Reputation Model

Hypothesis tests in section four led us to accept our alternative hypotheses that subjects do form reputations as 'trustworthy' types. In this section we compute maximum likelihood estimates of the proportion of 'trusting' and 'trustworthy' types. These estimates can be used to improve our understanding of the mechanism for reputation building behavior. A 'trustworthy' type will reciprocate in period 2, i.e., $3m_2-k_2>m_2$. From the date we observe n=19 cases where $m_2>0$. Of these r=7 cases reciprocated. Given our subjects are randomly sampled from our subject pool, and the double blind conditions, we assume that decisions are independent observations and that the number of successes, i.e., tpes, are given by the binomial distribution B(n,p). The likelihood function is given by, and the mazimum likelihood estimate of p is 0.37.

$$L(p|7,12)=Cp^{7}(1-p)^{12},$$

Table 2

Statistical Comparisons Of Treatment Effects
(KW = Kruskal-Wallis Statistic, Z = Normalized Wilcozon z)
(P = Probability of Accepting the Null Hypothesis of No Effect)

A. Comparisons of Amounts Sent

	Period 1 (N = 23)	Period 2 (N = 23)
No History (N = 32)	KW = 3.36 (P = 0.07)	KW = 0.81 (P = 0.37)
Period 1 (N = 12)		Z = 83.5 (P = 0.87)

B. Comparisons of Percentage Returns

	Period 1 (N = 23)	Period 2 (N = 19)
No History (N = 32)	KW = 10.29 (P = 0.002)	KW = 0.76 (P = 0.38)
Period 1 (N = 12)		Z = 16.00 (P = 0.0004)**

**These treatment comparisons are significant at 5%.

The 95% confidence interval for our estimated p is (.19,.59). See Amemiya (1994) pp 162-163.

A 'trusting' type will send m2≥ m1 as long as 3m1-k1>m1. From the data we observe n=20 cases that satisfy this restriction. Of these cases s=14 sent at least as much in the second period as they did in the first. Assuming that the number of successes is give by the Binomial distribution, the likelihood function is given by,

$$L(q|14,6)=Cp^{14}(1-q)^{6},$$

And the maximum liklehood estimate of q is .7. Note this is above the threshold of ½. The 95% confidence interval for our estimated q is (.48, .86).

These estimates suggest that it is the high incidence of 'trusting' types which makes it in the interest of room B subjects to act 'trustworthy.' Furthrmore, our high estimate of 'trusting' types helps explain why we do not get strongly significant differences in amounts sent between periods, while our lower estimate of 'trustworthy' types helps to explain the significant difference in returns.

5. Conclusions

That No History baseline provides evidence of reciprocity. To test if reciprocity can be explained by the existence of 'trusting' and 'trustworthy' types we design an experiment to see if subjects take into account these types. From this experiment we conclude that interpersonal history leads to a decision by room B subjects to build reputations as 'trustworthy' types. Using the No History baseline as a point of reference for one-shot levels of trust and reciprocity we find that the level of reciprocity is significantly higher in period one, but the level returns to the one-shot standard in period two. Furthermore directly comparing the levels of reciprocity between periods also leads us to conclude that reciprocity is significantly higher in period one. Subsequent estimation of types suggests that building a reputation as a 'trustworthy' is rational given the high proportion of 'trusting' room A subjects. This form of trust is consistent with the

definition of calculative trust defined in Williamson (1993).

One objections to these results is that subjects did not understand the investment game until the second stage, and differences were found due to this fact alone. To examine this concern twelve of the pairs run in theis experiment had participated in a pervious one=shot trust experiment using similar procedures. We find no significant differences between inexperienced and experienced subject's decisions. However, it is interesting to note that all three of our defections in period one occur with experienced subjects. Yet, when we control for experience we find that experienced and inexperienced groups both exhibit significant reputation effects.

Our high estimates of the proportion of 'trusting' types is consistent with the findings of McKelvey and Palfrey (1992) and Camerer and Weigelt (1988). In these experiments subjects behavior could be explained by a home grown priors on beliefs about the cooperativeness of their counterpart's population. These studies estimate beliefs which are higher than the actual level of cooperativeness. Why are beliefs higher? One possibility is that subjects actual field experience has led them to adopt a strategy where to first try cooperation but to quickly change their behavior if they learn that they are not dealing with a cooperative type. For example, in the two period data all three times a room B subject returned less than was sent their room A counterpart sent nothing the second period.

Of particular interest in the data was that reciprocity in period one was tightly distributed around sending back half of the tripled amount received. This is not the case in the baseline or in period two. This suggest that room B subjects may have been concerned with room A's inferences about their trustworthiness. If this is the case, then a 'fairness' norm may exist for the detection of cheaters. See Hoffman, McCabe, Smith (1995) for a more extensive discussion of the role of cheater detection in supporting reciprocity.

We also observed that whenever subjects end $10 in both periods, the total return was never non-negative. A similar, somewhat weaker, result is found in the one-shot games studied in BDM. Here are twelve cases where $10 is sent, eight exhibit non-negative return. These results

provides support for Robert Frank's hypothesis that trust is mutual. Sending $10 in one shot games or $20 in the two period game is a strong signal that a room A player is trying to use trust to make both players better off. The fact that this strategy seems to work reasonably well suggests that many more people (than the proportions observed in our experiments) are willing to reciprocate when they are dealing with a mutual trust relation, but small deviations from complete trust (for example sending less than $10) make them wary that such a relation exists.

REFERENCES

Arrow, Kenneth (1974), The Limits Of Organization, New York: Norton Press, York.

Bergm J., Dickhaut,J., & McCabe,K. (1994), "Trust, Reciprocity, and Social History," forthcoming in Games And Economic Behavior.

Camerer, C. And K. Weigelt (1998), "Experimental Tests of a Sequential Equilibrium Reputations Model," Econometrica, 56, 1-36.

Cosmides L. and J. Tooby 91992), "cognitive Adaptations for Social Exchange," in The Adapted Mind, edited by J. Barkow,L. Cosmides, and J. Tooby, New York: Oxford University Press.

Coleman J. (1990), Foundations of Social Choice Theory, Cambridge Massachusetts: Harvard University Press.

Fehr, E., G. Kirchsteiger, and A. Reidl (1993), "Does Fairness Prevent Market Clearing? An Experimental Investigation," Quarterly Journal Of Economics, 437-459.

Forsythe, R., J. Horowitz,N.Savin, and M.Sefton (1994), "Reliability, Fairness and Pay In Experiments with Simple Bargaining Games," Games And Economic Behavior, 6, 347-369.

Frank, Robert, Passions Within Reason: The Strategic Role of the Emotions, W.W. Norton and Company Inc, New York, 1988.

Harrison, g., and K. McCabe (1993), "the role of experience for testing bargaining theory in experiments, " Research In Experimental Economics, vol. 5, JAI Press, 137-169.

Hoffman E., K. McCabe, and V. Smith (1994), "Preferences, Property Rights and Anonymity in Bargaining Games," Games and Economic Behavior, 7 346-380.

Hoffman, E, K. McCabe, and V. Smith (1995), "Social Distance and Other Regarding Behavior in Dictator Games, " forthcoming in America Economic Review.

Kreps, David M. (1990), "Corporate Culture and Economic Theory," in Perspectives On Positive Political Economy, James alt and Kenneth Shepsle eds., Cambridge Universiy Press, Cambridge.

McKelvey, R., and T. Palfrey, (1992) "An Experimental Study Of The Centipede Game," Econometrica, 60, 803-836.

Rabin, M. (1993), "Incorporating Fairness into Game Theory and Economics," American Economic Review, 83, 1281-1302.

Van Huyck, J., R.Battalio, and M. Walters, (1993) "Commitment Verses Discretion in the Peasant-Dictator Game: Aggregate Analysis," mimeo, Texas A&M Economics.

Williamson, O. (1993), "Calculativeness, Trust, and Economic Organization, "Journal of Law and Economics, 36, 453-486.

Appendix A: Instructions For Trust Experiment

INSTRUCTIONS FOR ROOM A

You have been asked to participate in an economics experiment. The instructions you are about to read are self explanatory. We will not answer any questions during this experiment. If you have any questions, you should read back through these instructions. Now that the experiment has begun, we ask that you do not talk, at all, during this experiment. You will keep these instructions for the entire experiment.

In this experiment each of you will be paired with a different person who is in another room. You will not be told who these people are either during or after the experiment. This is room A other participants are in room B.

You will notice that there are other people in the same room with you who are also participating in this experiment. You will not be paired with any of these people. A person in room A, called Monitor A, and a person in room B, called Monitor B, will be chosen for today's experiment. The monitors will be in charge of the envelopes as explained below. In addition, the monitors will check that these instructions have been followed as they appear here.

Each person in room A and each person in room B has been given $20 as a show up fee for this experiment. At two different times, persons in room A will send in an envelope, up to $10 of their show up fee to a person in room B. Each dollar sent to room B will be tripled. For example, if you send an envelope which contains $2, the envelope will contain $6 when it reaches room B. If you send an envelope which contains $9, the envelope will contain $27 when it reaches room B. Each time that the money has been sent from room A, the counterpart in room B will then decide how much of the triple money to send back to the person in room A.

The experiment, therefore, consists of two rounds. In round one, people in room A will send up to $10 to a paired person in room B. This money will be tripled in transit. Each person in room B will then return none, some, or all of the money they received. Round two will repeat round one. The person that you are paired with will be the same for both rounds.

The remainder of these instructions will explain exactly how this experiment is run. This experiment is structured so that no one, including the experimenters and monitors, will know the personal decision of the people in either room A or room B. Since your decisions are private we ask that you do not tell anyone your decisions either during or after the experiment.

ROUND ONE

Room A people get their envelopes

The experiment is conduced as follows: Large unmarked envelopes have been placed in a box in room A. Each of these envelopes contains 10 one dollar bills (half of the show up fee for a person in room A), a smaller inner envelope, and a key in a sealed envelope marked Key. The inner envelope and key are marked with the same letter of the alphabet. The monitor, in room A, will point to one person at a time, and hand that person an unmarked envelope from the box. The person who was pointed to will then go to one of the seats with a partition and privately open the unmarked envelope. Only the person who opened the envelope will know which letter of the alphabet was in the envelope. Do not open the envelope market KEY until you are told to do so. The monitor will then point to the next person, and continue in this fashion until everyone has made their decisions. During this time, Monitor B will give each person in room B $10, half of their show up fee.

Room A people make their decision

Each person in room A must decide how many dollar bills to put in the inner envelope. The person then pockets the remaining dollar bills and the envelope marked KEY. Examples: (1) Put $2 in the inner envelope, and pocket $8 as well as the envelope marked KEY. (2) Put $9 in the inner envelope, and pocket $1 as well as the envelope market KEY. These are examples only, the actual decisions are up to each person.

Once a person in room A has made a decision they should put the inner envelope back inside the large unmarked envelope, and return the unmarked envelope to the box marked re-turn envelopes. Persons in room A should make sure that they have kept the envelope marked KEY as they will use this later in the experiment. Notice that each envelope returned win look exactly the same.

Monitor A transports envelopes to the recorder

After all the envelopes have been put in the return box Monitor A will transport the box to a recorder who is in the hallway. With Monitor A observing, the recorder will then, one at a time, take the inner envelope out of the unmarked envelope and then record on a blank sheet of paper, the letter of the envelope, and the amount of money in the envelope. While Monitor A is observing, the recorder will then triple the amount of money that was in the inner envelope and place the inner envelope back into the unmarked outer envelope. An envelope marked KEY will also be added to each outer envelope. At this point, the recorder will signal Monitor B to come to the recorder's desk Once Monitor B has arrived, Monitor A will be asked to return to room A

Room B people get their envelopes and make their decision

Monitor B will then carry the box of envelopes to room B. Monitor B will then point to one person at a time, and hand that person an unmarked envelope from the box. The person who was called will then go to a seat with a partition and then privately open the outer envelope. The person selected should first pocket the envelope marked KEY. Then, ach person in room B must decide how many dollar bills to leave in the inner envelope. The person then pockets the remaining dollar bills. The inner envelope should then be placed in the unmarked outer envelop and the outer envelope should then be placed in the box marked return envelopes.

Monitor B transports envelopes to the recorder

After all the envelopes in room B are returned, Monitor B will transport the box to the recorder in the hallway. The recorder will then, one at a time, open the inner envelope and record on a blank sheet of paper, the letter on the envelope, and then amount of money in the inner envelope. This is the second half of the show up fee for the people in room A. The recorder will then signal Monitor A to come to the recorder's desk. Once Monitor A has arrived, Monitor B will return to room B.

Recorder and Monitor A put envelopes in mailboxes

When Monitor A arrives the monitor and the recorder will carry the box of envelopes to room C directly opposite room A. Room C contains mailboxes with identifying letters. The letters correspond to the letters on the inner envelopes. While the recorder observes, Monitor A will place each envelope in the mailbox with the corresponding letter. All the mailboxes will then be locked. The recorder will then leave room C and monitor A will go to room A.

Room A people pick up their envelopes from mailboxes

Monitor A will then point to one person at a time from room A. That person will then enter room C alone and open the envelope marked KEY. Inside this envelope is a lettered key that will open the mailbox with the corresponding letter. The person from room A will then go to the appropriate mailbox, open it, take out the envelope, and remove the money in the outer envelope.

ROUND TWO

Room A people make their decisions

The money in the inner envelope is the $10 to be used in round two. In round two, you will be paired with the same person. The maximum amount that can be sent from room A is $10. The only money that can be sent from room B must have just delivered. While in the mailroom, each person should decide how much of the $10 to leave in the inner envelope to be sent to room B, and how much to keep themselves.

Once the person has made their decision, they should place the inner envelope in the larger one and return the envelope to the mailbox. They will then lock the mailbox and return to

room A. During this time the people in room B will each receive $10, the second half of their show up fee. Once the people in room A have finished their trips to the mailroom, the amounts in the inner envelopes will be recorded, tripled, and transferred to their matching mailboxes for the people in room B to pick up.

Room B people get their envelopes and make their decisions

Monitor B will then point to one person at a time to go to room C. The person pointed to will then go to room C and open their envelope marked KEY. The key has been marked with the letter of the mailbox that it opens. The person should then open their mailbox, and decide how much of the money to keep and how much to leave in the envelope to be returned to room A. They should then lock their mailbox and return the key to its envelope

Room B people leave

When leaving room C, they should return the envelope and key to the box marked keys. When people from room B are called to go to room C, they should take all of their belongings since they will be asked to leave the building when they are done.

Room A people pick up envelopes form mailboxes

The amounts in the inner envelopes will then be reordered and placed back in their mailboxes. The recorder and monitor B will then go to room B. Monitor A will then call out people in room A to go to room C. People should then open their mailbox and remove all of their money from the inner envelope.

Room A people leave

The person will then return the key to the envelope marked KEY and drop the envelope in the box just outside the door in the hallway. When you are called from room A to go to room C for the second time, you should take all of your belongings since you will be asked to leave the building when you are done.

When everyone in room A has left, the experiment is over, and the monitors will be paid for their participation.

Appendix B: Data From Experiments

Data From Experiments In Ordered Quartets

(Sent, Returned, Sent, Returned)

Subjects Who Participated In Previous Trust Experiments:

7/20a	7/20b	7/21
(10, 15, 10, 5)	(10, 15, 10, 15)	(10, 13, 10, 15)
(3, 5, 2, 0)	(1, 2, 2, 0)	(6, 3, 0, 0)
(2, 1, 0, 0)	(4, 7, 8, 4)	(10, 21, 10, 21)
(10, 15, 0, 0)		(10, 5, 0, 0)
(3, 5, 5, 0)		

Subjects Who Had Not Participated In Previous Trust Experiments

8/8a	8/8b	8/9
(5, 7, 10, 0)	(10, 15, 8, 9)	(4, 6, 7, 0)
(10, 30, 10, 2)	(3, 5, 4, 1)	(10, 20, 5, 12)
(8, 10, 8, 1)	(7, 12, 1, 0)	(10, 15, 10, 25)
	(5, 7, 6, 7)	