GAMES AND ECONOMIC BEHAVIOR **10,** 122–142 (1995)

Trust, Reciprocity, and Social History*

JOYCE BERG

University of Iowa, Iowa City, Iowa 52442

AND

JOHN DICKHAUT AND KEVIN MCCABE

Department of Accounting, University of Minnesota, 645 Management and Economics, 271 19th Avenue South, Minneapolis, Minnesota 55455-0413

Received September 15, 1994

We designed an experiment to study trust and reciprocity in an investment setting. This design controls for alternative explanations of behavior including repeat game reputation effects, contractual precommitments, and punishment threats. Observed decisions suggest that reciprocity exists as a basic element of human behavior and that this is accounted for in the trust extended to an anonymous counterpart. A second treatment, social history, identifies conditions which strengthen the relationship between trust and reciprocity. © 1995 Academic Press, Inc.

1. INTRODUCTION

A fundamental assumption in economics is that individuals act in their own self interest. To enable explanation and prediction this assumption is often combined with other assumptions: (1) the objects of interest are restricted to personal consumption; (2) more consumption is preferred to less; and (3) only current consumption possibilities and plans for future consumption can influence current behavior. In individual choice settings, behavior that deviates from self interest is viewed as irrational. However,

* The authors thank Richard Thaler for extensive conversations concerning the design of this experiment. John Carter, Werner Guth, Tom Palfrey, Reinhard Selten, the participants of the Second Amsterdam Workshop on Experimental Economics (AWEE '93), and two anonymous referees have provided valuable insights into our study.

0899-8256/95 $12.00

in group settings there exist situations where self-interested behavior makes everyone worse off. This paradox has led to the consideration of modes of behavior that are not so restrictive. In this paper we consider the role of trust in a two-person exchange.

In his lectures on the limits of organizations, Arrow (1974) notes that in the face of transaction costs trust is ubiquitous to almost every economic transaction. This observation raises important questions about economic behavior. Is trust a primitive in economic models of behavior? What factors increase (or decrease) the likelihood of trust in economic transactions? We provide answers to these questions in a specific experimental setting, the investment game. By guaranteeing complete anonymity and by having subjects play the investment game only once, we eliminate mechanisms which could sustain investment without trust; these mechanisms include reputations from repeat interactions, contractual precommitments, and potential punishment threats. We then show that positive investments still occur, suggesting that trust is an economic primitive.

The investment game is played as follows. Subjects in room A decide how much of their $10 show-up fee to send to an anonymous counterpart in room B. Subjects were informed that each dollar sent would triple by the time it reached room B. Subjects in room B then decide how much of the tripled money to keep and how much to send back to their respective counterparts. The unique Nash equilibrium prediction for this game, with perfect information, is to send zero money.[1] This prediction is rejected in our first (no history) treatment where 30 of 32 room A subjects sent money ($5.16 on average). How can we explain this behavior?

From a rational choice perspective, subjects who sent money must have believed their expected return would be positive; but given the noncooperative prediction, why would they believe this? One approach is to derive trust as an equilibrium of a repeated game or as reputation building in a

[1] The investment game is similar to the trust game in Kreps (1990), the centipede game in Rosenthal (1982), and the peasant–dictator game studied by Van Huyck, *et al.* (1993). The trust game has the same basic two-stage structure as the investment game, that is, passing money to the other player is risky but leads to an expanded pie, some of which may be returned. The centipede game may go on for many stages, but any two consecutive stages involve the same basic structure. 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 Palfrey (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 must cooperate to reach later stages. McKelvey and Palfrey find 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 investment game proposed in this paper provides a "boundary" design for the centipede game.

sequence of games; see for example Kreps (1990). Kreps notes that equilibria where trust emerges are but a small subset of the possible equilibria that may occur. While the theory of repeated games is useful for explaining how trust relationships become stable or unstable, it is not useful in discovering whether trust will exist absent repeat interactions, reputations, contracts, and punishment options.

A second approach is to derive trust as an evolutionarily stable strategy. Arrow (1974) refers to trust as an "important lubricant of a social system." While the emergence of trust is not formalized, he does categorize trust as a positive externality and suggests that evolutionary models may be able to explain the emergence of trust. The evolutionary approach is currently being studied in economics by Hirshleifer (1977), Guth and Kliment (1993), Guth *et al.* (1993), and Selten (1989) and in psychology by Cosmides and Tooby (1992). Evolutionary models predict the emergence of trust because it maximizes genetic fitness even though myopic self interest suggests cheating. From this perspective, trust can be viewed as a behavioral primitive that guides behavior in new situations; however, it may in some circumstances become extinct or it may be superseded by an individual's capacity to engage in self-interested decisionmaking.

Our "no history" results suggest that reciprocity does occur. We conjectured that the decision to reciprocate may depend on a counterpart's interpretation of room A behavior. If the counterpart interprets the decision to send money as an attempt to use trust to improve the outcome for both parties, then the counterpart is more likely to reciprocate. This lead to the consideration of factors that encourage the trust interpretation. One such factor is social history. There are many instances within organizations where public information reflects an organization's social history. Since social history provides common information about the use of trust within an organization, such a history may reinforce individuals' predispositions toward trust. This observation led to a second (social history) treatment. The following changes were made to the no history treatment: (1) subjects were recruited who had not participated in the previous sessions and (2) these subjects were given a summary of the no history results as part of their instructions.

2. Hypotheses and Procedures

The investment game is played as follows. In stage one, the subjects in rooms A and B are each given \$10 as a show-up fee. While subjects in room B pocket their show-up fees, subjects in room A must decide how much of their \$10 to send to an anonymous counterpart in room B. We denote this amount by M_a. The amount sent is then tripled, resulting in a

total return of $3M_a$. In stage two, a counterpart in room B is given the tripled money. The counterpart must decide how much money to return, which is denoted $k_b(3M_a)$. The room A subject chooses the strategy M_a in $\{0, 1, 2, ..., 10\}$, while the counterpart in room B chooses the strategy

$$k_b: \{0, 3, ..., 30\} \rightarrow \{0, 1, ..., 30\},$$

which satisfies $0 \leq k_b(3M_a) \leq 3M_a$.

These strategies result in the payoffs

$$P_a(M_a, k_b) = \$10 - M_a + k_b(3M_a),$$

and

$$P_b(M_a, k_b) = 3M_a - k_b(3M_a).$$

A subject's initial wealth is denoted W_i. If subjects have strictly increasing indirect utility function for wealth, given by $V_i(W_i + P_i(M_a, k_b))$ for i = a, b, and each subject, i, maximizes $V_i(\cdot)$, then subjects in room B have a dominant strategy to keep all the money, that is, $k_b(3M_a) = 0$ for all M_a. If room A subjects infer their counterpart's dominant strategy, then they should send nothing, i.e., $M_a = 0$.

Evidence from other experiments suggests that subjects do consider their counterparts move in two-stage sequential games.[2] One hypothesis for the investment game is that subjects will make decisions consistent with the subgame perfect prediction,

$$N_0: M_a = 0 \qquad \text{for all a.}$$

If for some reason a positive amount is sent, i.e., $M_a > 0$, then we can test for the dominant strategy property by room B subjects.

$$N_1: \text{If } M_a > 0, \text{ then } k_b(3M_a) = 0, \qquad \text{for all b.}$$

The investment game provides a role for the use of trust in achieving a joint improvement to the subgame perfect outcome. For purposes of this

[2] See for example the two- versus three-stage bargaining results in Neelin *et al.* (1988) and Harrison and McCabe (1993) and the mouselab studies of bargaining by Camerer, *et al.* (1991). See also the baseline results for extensive form play by McCabe, *et al.* (1994). The conclusion from this research is that subjects do look at least one stage ahead. This observation is sufficient for assuming that subjects in room A did consider their counterpart's dominant strategy in the two-stage investment game.

paper, our definition of trust is similar to Coleman's (1990, Chap. 5). In Coleman's terminology a room A subject is known as the "trustor"; sending $M_a > 0$ is said to "place a trust" in the trustee. The trustee is said to "keep the trust," i.e., reciprocate, when $k_b(3M_a) > M_a$. Trust can be defined in terms of the following two actions: first, the trustor gives a trustee the right to make a decision; and, second, the trustee makes a decision which affects both trustor and trustee. Subjects have used trust to facilitate exchange, if the following conditions are met. (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.[3]

The conditions for trust can be satisfied in the investment game as follows: (1) Sending money in stage one is risky since the counterpart may or may not reciprocate; (2) in stage two, the counterpart has to give up money to make the room A subject better off; and (3) since the money triples both parties can be made better off relative to the subgame perfect outcome. The existence of trust and reciprocity in these experiments support the following alternative hypotheses,

$$A_0: M_a > 0 \qquad \text{for some a,}$$

and

$$A_1: k_b(3M_a) > M_a.$$

From a learning or evolutionary games standpoint, a strategy will become extinct if its relative payoff is poor. Since the simple strategy of not investing results in zero profit, trust cannot emerge as a norm unless it satisfies the positive profits hypothesis that, for at least some amounts M_a, the average net return is positive; i.e., if there are N room B subjects each receiving the total amount $3M_a$, then

A_2: There exist amounts M_a such that

$$\sum_b \frac{k_b(3M_a)}{N} > M_a.$$

[3] Williamson (1993) uses the same basic definition of trust in his review of the trust literature, but he distinguishes between calculative trust and personal trust. Calculative trust occurs when the trustor decides whether or not to entrust the trustee on the basis of an expected utility calculation. Personal trust is based on more of a noncalculative "feeling" that entrusting the trustee is the correct decision. In summary, trust is understood as a reliance on the reciprocity of others where a return is made for something done or given. Different forms of trust emerge depending on the basis of the reliance.

An important feature of the investment game is that different amounts M_a can be sent from room A. It seems plausible that the amount sent may have an impact on reciprocity. At one extreme, sending \$1 may signal a very weak belief in reciprocity; at the other extreme, sending \$10 may signal a strong belief in reciprocity. From an evolutionary perspective, someone with a predisposition to reciprocate may be more willing to reciprocate when they believe their counterpart shares a common regard for trust. Rabin (1993) makes such an assumption by incorporating a "kindness" function into subjects' utility in such a way as to capture the following behavior: as one's counterpart increases his or her "kindness," the utility maximizing response is to be kinder in return. These observations suggest the following hypothesis.

$$A_3: M_a \text{ and } k_b(M_a) \text{ are positively correlated.}$$

A. *The Double Blind Procedure*

Stage two of the investment game is a dictator game[4], except that now subjects in room B, acting as dictators, have been entrusted with the money by a counterpart in room A. Dictator games without this element of trust have been studied extensively by experimentalists in order to understand bargaining behavior. Forsythe *et al.* (1994; henceforth FHSS) ran a \$10 dictator game where the amount given by the subject dictators is higher than the self-interested prediction of zero. The FHSS results were replicated by Hoffman *et al.* (1994; henceforth HMSS), who further find that self-interested behavior is significantly increased by using a procedure (double-blind) which guarantees complete anonymity with respect to other subjects and the experimenter.

In implementing the investment game, we incorporated the double-blind procedures used by HMSS. Our choice was motivated by the desire to control for the possibility of repeat game effects and thus protect any observed results from being attributed to reputation, collusion, or the threat of punishment. Such phenomena become impossible when no one, including the experimenter, is able to map individual decisions back to the identity of the decision maker. Yet it must be possible for the experimenter to gather data.

[4] In a dictator game player one, called the dictator, is given a sum of money, often \$10. The dictator must then decide how much if any to send player two. A theory which assumes self-interested, non-satiated behavior predicts that the dictator will keep all the money.

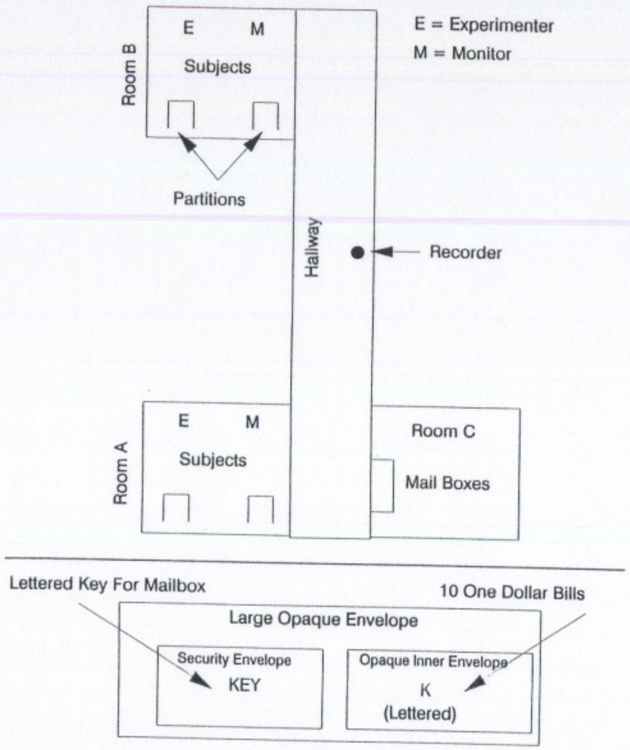

FIG. 1. Diagram of trust experiment.

Figure 1 shows the physical implementation of the investment game. This hand-run implementation involves passing envelopes among three rooms. Rooms A and B contain subjects, while room C contains a bank of 14 mailboxes. The experiment begins by having an experimenter (E) in each room read aloud the instructions that appear in appendix A. Once the instructions are finished, subjects in room A are called one at a time by the monitor (M). Once called, a subject receives a large unmarked opaque envelope which contains: (1) an inner envelope with a lettered mailbox key, (2) a smaller lettered opaque envelope which will contain the money to be invested, and (3) the subject's show-up fee of 10 one-dollar bills. The letter on the smaller envelope corresponds to the letter on the key, thus enabling subjects in room A to receive the return from their own investment decision.

Upon receiving the envelope a subject then proceeds to the back of the room, to a privacy preserving partition, and decides (1) how many dollars to send, i.e., put in the smaller opaque envelope which is then put back

into the larger opaque envelope, and (2) how many dollars to keep along with the lettered mailbox key. Upon completing the task, the subject drops the large opaque envelope into a return box and returns to his or her seat. Once all decisions in room A are made, the monitor from room A takes the envelopes to the recorder (R) in the hallway. At this time the amounts sent are recorded by subject letter and the amount invested is tripled and returned to the lettered envelope, which is placed back into the unmarked envelope.

When the recording is finished, the monitor from room A returns to room A and the monitor from room B is called to get the envelopes. In room B subjects are called up one at a time and given one of the large unmarked opaque envelopes. Subjects then go to the back of the room where they privately open the outer and inner envelopes and learn how many dollars are in the inner envelope. The subject then decides the amount he or she wants to pay back, by putting that amount of money back in the inner envelope, and keeps the remaining dollars. The inner envelope is then returned to the unmarked outer envelope and is placed in the return box. At this point the subject in room B exits the experiment.

Once all decisions in room B are made, the monitor from room B takes the envelopes back to the recorder in the hallway. At this time the money paid back to subjects in room A is recorded by the letter on the inner envelope. Once recording is done, the inner envelopes are then placed in their corresponding mailboxes in room C. At this point the monitor from room B and the recorder both return to room B.

The monitor from room A then calls one subject at a time to go to room C and privately open his or her mailbox. The subject then pockets the money in the inner envelope, returns the key to a drop box, and exits the experiment. At no time does anyone else learn what key or mailbox was used. This procedure continues until all subjects in room A are finished.

B. *Subjects*

Subjects were recruited from the undergraduate student population at the University of Minnesota. All subjects were recruited by phone and had previously participated in at least one unrelated paid laboratory experiment. Subjects received $10 for showing up and $3 if they were bumped due to overbooking. Subjects were told to report directly to either room A or room B. Once an equal number of subjects were in each room, monitors were chosen from each room by having subjects draw without replacement

TABLE I
EXPERIMENTAL SESSIONS

Treatment	Date	Number of pairs
No history	7/13/93	10
	7/19/93	10
	7/23/93	12
Social history	7/27/93	10
	7/31/93	9
	8/03/93	9

a chip from a cup until the chip marked M appeared. Experiments lasted from 60 to 90 min. Table I summarizes our sessions.

3. NO HISTORY RESULTS

Our no history treatment consisted of 32 pairs of subjects run over 3 days. Figure 2 graphs the resulting data from these sessions. The data are presented in descending order using the amount sent from room A,

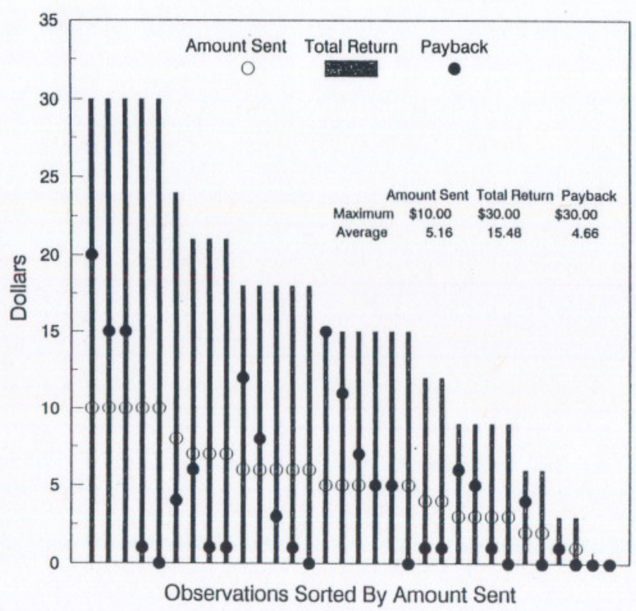

FIG. 2. Trust experiment results showing amount sent (○), total return (■), and payback (●). No history was provided to the subjects.

shown as an open circle, as the primary key. For example, 5 of 32 subjects in room A sent their entire show up fee of $10. Gray bars indicate the amount received by the counterpart in room B. The data are further sorted in descending order using the payback, the amount sent back to room A (shown as a black circle), as the secondary key. On average, $5.16 is sent by room A, resulting in an average payback of $4.66 by room B.

A. *Room A Decisions*

We make the following observations about room A decisions. (1) Only 2 of 32 subjects sent zero. (2) The amounts sent are highly variable, ranging from $1 to $10. These observations provide very little support for Hypothesis N_0. Furthermore, we cannot reject the null hypothesis that our sample of room A decisions was drawn from a uniform distribution over the amounts $\{0, 1, 2, \ldots, 10\}$. Consider the following approximate randomization test; see Edington (1987). Randomly assign the ($N = 32$) decisions to the possible amounts sent, $m = 0, 1, 2, \ldots, 10$, under the assumption that each amount is equally likely. The resulting sample can be denoted s_i and the frequency of each amount, m, can be denoted f_m. We then measure the variation in the sample, s_i as

$$\nu(s_n) = \sum_{m=0}^{11} \left(f_m - \frac{N}{11}\right)^2.$$

Given our actual sample denoted d, we can now ask what is the probability, p, that $\nu(s_i)$ is greater than or equal to $\nu(d)$. We then draw 100,000 random samples to calculate p to be 0.29.

B. *Room B Decisions*

We make the following observations about room B decisions (1) Of the 28 room B subjects who were sent $M_a > \$1$, 12 returned $0 or $1 to their counterparts. This behavior is supportive of hypothesis N_1. (2) However, 11 of the same 28 subjects returned more than their counterpart sent, resulting in positive net returns. This behavior is supportive of hypothesis A_1. (3) Investments of $5 had an average payback of $7.17, while investments of $10.00 had an average payback of $10.20. This observation provides some support for hypothesis A_2.

To examine hypothesis A_3 we calculate Spearman's rank correlation coefficient, r_s, between the paired room A and room B decisions. Since absolute amounts sent and amounts returned will bias our correlation statistic upwards, i.e., low amounts sent preclude some high returns, we compare amount sent to percentage returned. An $r_s = 0.01$ suggests no correlation between amounts sent and payback decisions.

4. Social History Results

The decisions to send money and reciprocate, in the no history treatment, raises the possibility that subjects decisions in the investment game are influenced by social norms.[5] Coleman (1990) defines a norm on a specific action to be a socially defined right by others to control an individual's action. Such control can be achieved through the use of sanctions or rewards. In Coleman's terminology there is a prescriptive norm to reinforce the room B decision to keep a trust, i.e., reciprocate. In light of this norm it is rational for individuals in room A to place a trust, i.e., send money.

In cases where outside sanctions or rewards are not available, norms may still exist if sanctions or rewards have been internalized by the decision maker. Coleman argues that norms are more likely to be internalized when an individual clearly identifies with a particular group. The process of having an individual identify with a group is termed socialization. In the no history experiment subjects were all University of Minnesota undergraduates. From this viewpoint, social history provides common information about the use of trust within a group; such a history may increase social identity and reinforce an individual's predisposition toward trust. By providing social history in a double-blind, one-shot setting, we focus on the internalization of social norms, as opposed to other potential mechanisms for reciprocity such as reputation building.

In the social history treatment 28 additional pairs of subjects were run over three days. Each subject was given a report summarizing the decisions of the previous 32 pairs of subjects in our no history treatment. This report appears at the end of the instructions in Appendix A. For each level, \$0–10, the report details the number of times that amount was sent, the average

[5] Coleman notes that demand for norms increases with the existence of externalities between subjects. In the investment game such an externality exists since money triples only if it is sent, while return on investment depends on payback decisions. But why do we only observe positive average returns on \$5 and \$10. Elster notes that norms often have an all-or-nothing feature. In the investment game room B subjects are more likely to reciprocate when room A subjects send "all" of their money. Note that all may be interpreted as \$10, but it could also be interpreted as \$5 due to our experimental design. Subjects had all previously participated in individual choice experiments where they were guaranteed a minimum of \$5 (for showing up) which could not be used in the experiment. From past experience some subjects may have split their \$10 into the \$5 guarantee and \$5 for potential investment decisions.

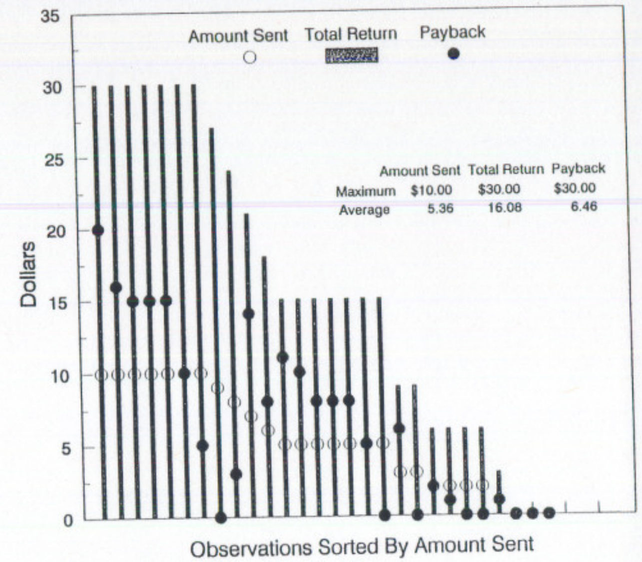

Fig. 3. Trust experiment results showing amount sent (○), total return (■), and payback (●). A social history was provided to the subjects.

payback (average amount returned), and the average net return (average amount returned minus average amount sent). In order to minimize any additional presentation effects in the instructions we referred to the report only in the following paragraph;

> "Each of you has received a report summarizing the decisions of the previous 32 pairs of subjects who have participated in this experiment in July. Please check the last page of the instructions to be sure you have this sheet."

If, in the no history treatment, subjects in room A failed to appreciate the dominant strategy property of the room B decisions, then providing a social history to room A makes subjects more aware of the existence of subjects in room B who do not reciprocate. This could result in a loss of trust, resulting in lower amounts sent. Such behavior should increase support for hypothesis N_0. Alternatively, room A subjects may focus on the positive net returns for the \$5 and \$10 levels. This focus could result in an increase in trust, resulting in more decisions to send either \$5 or \$10, providing increased support for hypothesis A_0. For room B subjects, social history may make keeping the money more acceptable, providing greater support for hypothesis N_1. Alternatively, social history may stimulate reciprocity when \$5 and \$10 are sent, providing support for hypotheses A_1–A_3.

Figure 3 graphs the results from these sessions. Data are sorted using

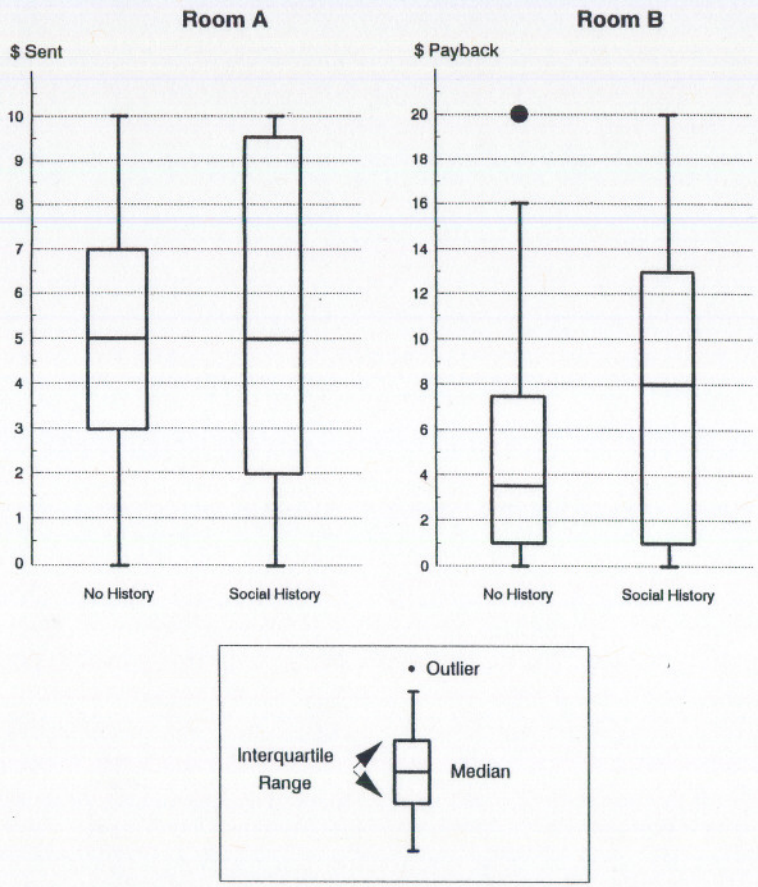

FIG. 4. Subjects' decision by room and treatment.

the same keys, amount sent and payback, as were used in Fig. 2. On the average room A subjects sent \$5.36, resulting in an average payback of \$6.46. Comparing paybacks under no history to paybacks under social history, a one-sided Wilcoxon rank–sum test ($r = 776$) is significant at the $p = 0.1$ level.

In Fig. 4 we summarize the data with box plots. These plots show a measure of location (the median) as a solid line, a measure of dispersion (the interquartile range) as a box and outliers (points) which are further than 1.5 times the interquartile range from the upper and lower quartiles. This range is shown as a vertical line which is drawn to the most extreme data point within 1.5 times the interquartile range from the upper and lower quartile boundaries. An immediate feature of the data is an increase

in dispersion of the data, i.e., the interquartile range, in the social history treatment. This increase explains the reduced significance of statistical tests comparing no history with social history.

A. *Room A Decisions*

We make the following observations about room A decisions: (1) Only 3 of 28 subjects sent zero. This observation provides very little support for hypothesis N_0. (2) \$5 and \$10 are now sent 50% of the time. However, this increase in frequency of \$5 and \$10 amounts is not statistically significant. In the table included in Appendix B we compare average net return by amount sent. Using the same randomization test, explained in Section 3, we now reject (at $p = .06$) the null hypothesis that our sample of room A decisions was drawn from a uniform distribution over the amounts {0, 1, 2, . . . , 10}.

B. *Room B Decisions*

We make the following observations about room B decisions. (1) Of the 24 room B subjects who were sent $M_a >$ \$1, 6 returned \$0 or \$1 to their counterparts; (2) 13 of the same 24 subjects returned more than their counterpart, sent resulting in positive net returns; (3) investments of \$5 had an average payback of \$7.14, while investments of \$10.00 had an average payback of \$13.17.

To examine Hypothesis $A3$ we calculate Spearman's rank correlation coefficient, r_s, between the paired room A and room B decisions. An $r_s = .34$ suggests an increase in correlation between amounts sent and payback decisions when social history is provided. To test if the change in correlation between no history and social history is significant we use resampling. We generate 5000 sample points of the following form. First, we randomly assign our 60 paired observations between the two treatments with 32 pairs and 28 pairs; we then compute the difference in correlation statistics for our sample. From this empirical distribution we compute the probability, $p = .06$, that our resampled differences in correlation statistics are greater than or equal to our actual difference. We conclude that the change was significant.

5. Payback Decisions and Earnings

In Fig. 5 we graph the joint earnings of the 60 (room A, room B) pairs. The large outer triangle with points {(0,40), (10,10), and (30,10)} indicates the set of feasible earnings pairs. The shaded triangle {(10,30), (10,10), (30,10)} indicates the earnings pairs with non-negative net returns to room

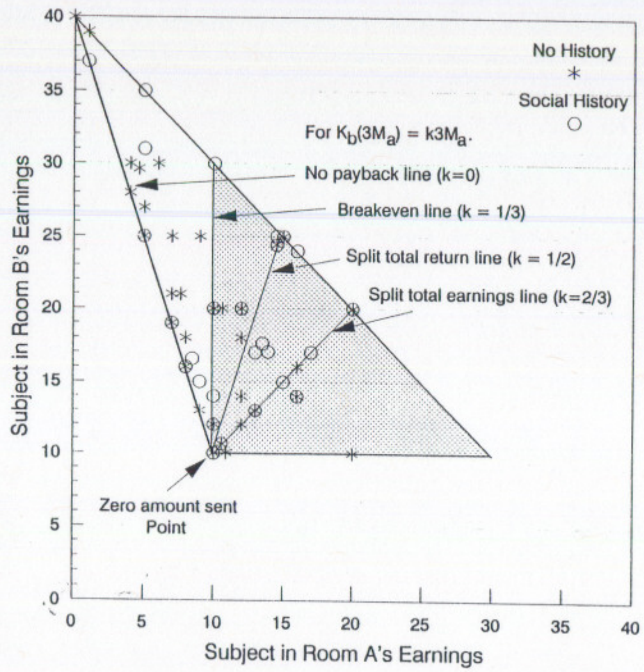

FIG. 5. Distribution of joint dollar earnings.

A, while the unshaded triangle {(0,40), (10,10), (10,30)} indicates the nonpositive net returns to room A. If we write the amount returned as a fraction of the total return, i.e., $k_b(3M_a) = k3M_a$, then we can parameterize the following line segments: (1) When $k = 0$, subject b keeps all the money, i.e., the line segment with endpoints (0,40) and (10,10); (2) when $k = 1/3$, subject b returns the original amount sent, allowing subject a to break even, i.e., the line segment (10,10) with endpoints (10,10) and (10,30); (3) when $k = 1/2$, subject b splits the total return, $3M_a/2$, with subject a, i.e., the line segment with endpoints (10,10) and (15,25); (4) when $k = 2/3$, subject b splits the net return with subject a, i.e., the line segment with end points (10,10) and (20,20) which splits total earnings inclusive of show-up fees.

We have plotted the 32 no history pairs as asterisks and the 28 social history pairs as open circles. When points coincide we have offset them slightly. Using k as a measure of room B types and allowing the joint earnings pair to be wtihin a dollar of the type k prediction, then the four types $k = 0$, $k = 1/3$, $k = 1/2$, and $k = 2/3$ explain over 90% of the data. While the data suggest that social history increases the $k = 1/3$ and greater types, the increase is not statistically significant.

6. CONCLUSIONS

We have identified an experimental design where trust can be used for mutual gain. The double-blind and one-shot controls used in this design strengthen our conclusion that self interest alone cannot explain our results. In particular, these controls rule out reputations or contractual precommitments as explanations of subjects' behavior. In our initial no history treatment 30 of 32 room A subjects sent money, while 11 of these 30 decisions resulted in a payback greater than the amount sent. One explanation of the data is that room A subjects were willing to place a trust, by risking some amount of money, in the belief that there would be reciprocity; the room B subjects who reciprocated kept this trust. The remaining two-thirds may have been acted out of self interest, or alternatively they may not have interpreted the room A decisions as placing a trust and thus reciprocation was not an issue.

The social history treatment explores the interpretation that some subjects reciprocated because they were trusted. Given that two-thirds of the room B subjects did not reciprocate in the no history treatment, it seemed plausible that providing this information would cause room A subjects to send less and justify room B decisions to reciprocate less. This was not the case. Moving from no history to social history resulted in three observable effects. First, there is a shift in average return from a negative $0.50 (no history) to a positive $1.10 (social history). Since the average amount sent only increased by 20 cents, most of this change is accounted for by an increase in the average payback of $1.80. Second, room A decisions become more systematic with respect to our randomization test. Our randomization results led us to the following direct comparison of behavior. Third, we observe an increased correlation between amounts sent and payback decisions.

Taken together our two treatments provide a strong rejection of the subgame perfect prediction that room A subjects will send no money. They did so 55 out of 60 times. Furthermore, we see reciprocation, but we also see room B types who did not reciprocate. Here the evidence is mixed. Some of the room B subjects who did not reciprocate may not have interpreted room A behavior as initiating a trust. The evidence on $5 and $10 amounts sent suggests that it may have been easier to interpret this as trust and thus served as focal points for trust. Further evidence comes from the social history treatment which was designed to increase subjects' propensity to believe in and use trust.

Trust can be used to explain a number of other experiments. In HMSS, 77 subjects in the double blind dictator games (DB1 and DB2), who received $10 from the experimenter, sent an average of 11 cents for every dollar received; in contrast, the 55 room B subjects, who were sent money in both treatments of the investment game, sent back an average of 35 cents for every dollar received. This suggests that forward signaling, i.e., sending money in the investment game, may be essential for reciprocity. When subjects are simply endowed as a benevolent dictator by the experimenter (as in HMSS) there is no call for reciprocation. Reciprocity carries over into markets as well. Fehr *et al.* (1993) look at the effect of paying higher wages on subjects productivity where moral hazard exists. They find that subjects who are paid more are more likely to reciprocate by shirking less. McKelvey and Palfrey (1992) find that subjects in centipede games are more likely to pass, thus allowing the game to continue even though the subgame perfect prediction is to end the game. They explain their results in terms of a reputation model with altruistic types, i.e., those who pass no matter what. An alternative interpretation of their data is the existence of types who will keep a trust where the round-one decision to pass places a trust which can be kept in round two. The trust can be placed again in round three, and so on. However, consistent with the finding of payoff variability in the investment game, once sufficient reciprocity is achieved to make both players better off either player is likely to end the game. Given the existence of subjects who keep a trust, as much as one-third of the subjects in our no history experiment, the other subjects in the centipede game have an incentive to act as one of these types.

The ability to include trust and reciprocity as part of the rational choice paradigm would seem to allow better explanations of economic institutions. For example, there is strong evidence that (1) "framing" joint decision problems in terms of property rights and (2) implementing procedures for assigning such rights changes subject behavior; see for example HMSS or Hoffman and Spitzer (1982, 1985). But how do these changes in language and procedures affect behavior? Based on our study, one interpretation is that the language and institution of property rights is built on the behavioral primitives trust and reciprocity. Thus, by inventing property rights and allowing social history, society stimulates norms of behavior which further strengthen reciprocity. In earlier experiments property rights protect first movers from punishment, a negative form of reciprocity, when first movers make decisions which are more in their own self interest. By contrast, in the investment game, their is a property right when subject a sends part (or all) of subject a's show-up fee to room B, thus entrusting private property to another subject.

In conclusion, experiments on ultimatum game, repeated prisoners' dilemma games, and other extensive form games provide strong evidence that people do punish inappropriate behavior even though this is personally costly. Furthermore, subjects take this into account when they make their decisions. The investment game provides evidence that people are also willing to reward appropriate behavior and that this too is taken into account. Taken together, these results suggest that both positive and negative forms of reciprocity exist and must be taken into account in order to explain the development of institutional forms which reinforce the propensity to reciprocate. Our results provide strong support for current research efforts which attempt to measure trust, see Cummings *et al.* (1994), and efforts to integrate reciprocity into standard game theory, see Geanakopolos *et al.* (1989) and Rabin (1993).

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.

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 verify that the instructions have been followed as they appear here.

Each person in room A and each person in room B has been given $10 as a show up fee for this experiment. Persons in room A will have the opportunity to send in an envelope, some, all, or none 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. The person in room B will then decide how much money to send back to the person in room A and how much money to keep.

(For the social history treatment we added the following paragraph)
Each of you has received a report summarizing the decisions of the previous 32 pairs of subjects who have participated in this experiment during July. Please check the last page of the instructions to be sure you have this sheet.

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 people in either room A or room B. Since your decision is private we ask that you do not tell anyone your decision either during, or after, the experiment.

The experiment is conducted as follows: Twelve large unmarked envelopes have been placed in a box in room A. Each of these envelopes contains 10 one dollar bills (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 large box on top and privately open the unmarked envelope inside the box. Only the person who opened the envelope will know which letter of the alphabet was in the envelope. Do not open the envelope marked 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.

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 marked KEY. These are examples only, the actual decision is 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 return 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 will look exactly the same.

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 record on a blank sheet of paper, the letter on the envelope, and the amount of money in the envelope. While monitor A is observing, the recorder will then triple the amount of money in the inner envelope, and place the inner envelope back into the unmarked outer envelope. At this point, the recorder will signal the monitor from room B to come to the recorders desk. Once monitor B has arrived monitor A will be asked to return to room A.

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 large box on top and then privately open the outer envelope inside the box. The monitor will then point to the next person. Each 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 envelope and the outer envelope should then be placed in the box marked return envelopes. The person in room B will then be asked to leave since the experiment is over for that person. When you leave we ask that you leave the building. 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 the amount of money in the inner envelope. The recorder will then signal monitor A to come to the recorders desk. Once monitor A has arrived monitor B will return to room B.

When monitor A arrives the monitor and 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 inner envelope in the box with the corresponding letter. All the mailboxes will then be locked. The recorder will then go back to room B and monitor A will go to room A.

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 which will open the mailbox with the corresponding letter. The inner envelope in the mailbox is the same one the person in room A started with. We have underlined the letters H and I on the key tags to make it clear which letter you have. The person from room A will then go to the appropriate mailbox, open it, take out the envelope, and remove the money. The person will then return the envelope to the mailbox and lock the mailbox. 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 to go to room C you should take all 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.

(Report included in Social History Treatment)
Summary of Previous Choices Aggregated by Amount Sent

Amount sent	0.00	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00
Number of persons Choosing this amount	2	2	2	4	2	6	5	3	1	0	5
Average amount returned	0.00	.50	2.00	3.00	1.00	7.17	4.80	2.67	4.00		10.20
Amount sent	0.00	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00		10.00
Average profit		−.50	0.00	0.00	−3.00	2.17	−1.20	−4.33	−4.00		0.20

APPENDIX B: DATA FROM EXPERIMENTS IN ORDERED PAIRS

(Investment, Payback)

	No history			Social history		
	7/13/92	7/19/92	7/23/92	7/27/92	7/31/92	8/03/92
1	(7,1)	(5,7)	(6,12)	(5,11)	(3,0)	(1,1)
2	(3,0)	(10,0)	(10,15)	(10,15)	(2,0)	(10,20)
3	(7,6)	(5,0)	(3,6)	(2,2)	(10,5)	(7,14)
4	(5,11)	(8,4)	(6,8)	(5,8)	(2,0)	(10,15)
5	(3,1)	(5,15)	(6,1)	(0,0)	(5,0)	(3,6)
6	(2,4)	(0,0)	(10,15)	(2,1)	(5,10)	(10,10)
7	(6,0)	(7,1)	(10,1)	(10,16)	(8,3)	(6,8)
8	(4,1)	(1,0)	(6,3)	(10,15)	(5,8)	(5,8)
9	(10,20)	(3,5)	(4,1)	(5,5)	(0,0)	(0,0)
10	(5,5)	(2,0)	(0,0)	(9,0)		
11			(5,5)			
12			(1,1)			

REFERENCES

ARROW, K. (1974). *The Limits Of Organization.* New York: Norton, York.

CAMERER, C., JOHNSON, E., RYMON, T., AND SEN, S. (1993). "Cognition and Framing in Sequential Bargaining for Gains and Losses," in *Frontiers of Game Theory*, (K. Binmore, A. Kirman, and P. Tuni, (Eds.), Cambridge, MA: MIT Press.

COSMIDES, L., AND TOOBY, J. (1992). "Cognitive Adaptations for Social Exchange," in *The Adapted Mind*, (J. Barkow, L. Cosmides, and J. Tooby, Eds.). New York: Oxford Univ. Press.

COLEMAN, J. (1990). *Foundations of Social Choice Theory.* Cambridge, MA: Harvard Univ. Press.

CUMMINGS, L., AND BROMILEY, P. (1995). "The Organizational Trust Inventory (OTI): Development and Validation," in *Trust in Organizations* (R. Kramer and T. Tyler, Eds.), Twin Oaks, CA: Sage.

FEHR, E., KIRCHSTEIGER, G., AND RIEDL, A. (1993). "Does Fairness Prevent Market Clearing? An Experimental Investigation," *Quart. J. Econ.* **108,** 437–459.

Forsythe, R., Horowitz, J., Savin, N., and Sefton, M. (1994). "Replicability, Fairness and Pay in Experiments with Simple Bargaining Games," *Games Econ. Behav.* **6,** 347–369.

Geanakoplos, J., Pearce, D., and Stachetti, E. (1989). "Psychological Games and Sequential Rationality," *Games Econ. Behav.* **1,** 60–79.

Guth, W., and Kliment, H. (1993). "Competition or Cooperation: On the Evolutionary Economics Of Trust, Exploitation and Moral Attitudes," mimeo, University of Frankfurt.

Guth, W., Ockenfels, P., and Wendel (1993). "Efficiency by Trust in Fairness? Multiperiod Ultimatum Bargaining Experiments with an Increasing Cake," *Int. J. Game Theory* **22,** 51–73.

Harrison, G., and McCabe, K. (1993). "The Role of Experience for Testing Bargaining Theory in Experiments," *Res. Exp. Econ.* **5,** 137–169.

Hirshleifer, J. (1977). "Economics from a Biological Viewpoint," *J. Law Econ.* **20,** 1–52.

Hoffman, E., McCabe, K., Shachat, K., and Smith, V. (1994). "Preferences, Property Rights, and Anonymity in Bargaining Games," *Games Econ. Behav.* **7,** 346–380.

Hoffman, E., and Spitzer, M. (1982). "The Coase Theorem: Some Experimental Tests," *J. Law Econ.* **25,** 73–98.

Hoffman, E., and Spitzer, M. (1985). "Entitlements, Rights, and Fairness: An Experimental Examination of Subjects' Concepts of Distributive Justice," *J. Legal Studies* **15,** 254–297.

Kreps, D. M. (1990). "Corporate Culture and Economic Theory," in *Perspectives on Positive Political Economy* (J. Alt and K. Shepsle, Eds.). Cambridge: Cambridge Univ. Press.

McCabe, K., Rassenti, S., and Smith, V. (1994). "Forward and Backward Rationality in the Achievement of Cooperation," mimeo, Economic Science Laboratory, University of Arizona.

McKelvey, R., and Palfrey, T. (1992). "An Experimental Study of the Centipede Game," *Econometrica* **60,** 803–836.

Neelin, Sonnenschein, and Spiegel (1988). "A further test of noncooperative bargaining theory: Comment," *Amer. Econ. Rev.* **78,** 824–836.

Rabin, M. (1993). "Incorporating Fairness into Game Theory and Economics," *Amer. Econ. Rev.* **83,** 1281–1302.

Rosenthal, R. (1982). "Games of Perfect Information, Predatory Pricing, and the Chain-Store Paradox," *J. Econ. Theory* **25,** 92–100.

Schelling, T. (1980). *The Strategy Of Conflict.* Cambridge, MA: Harvard Univ. Press.

Selten, R. (1989). "Evolution, Learning, and Economic Behavior," the 1989 Nancy L. Schwartz Memorial Lecture, Northwest University, 1989.

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

Williamson, O. (1993). "Calculativeness, Trust, and Economic Organization," *J. Law Econ.* **36,** 453–486.