ON THE NATURE OF RECIPROCAL MOTIVES

James C. Cox
Department of Economics
University of Arizona
jcox@eller.arizona.edu

Cary A. Deck
Department of Economics
University of Arkansas
cdeck@walton.uark.edu

September 2000; revised August 2002

ON THE NATURE OF RECIPROCAL MOTIVES*

By James C. Cox and Cary A. Deck

Data from eleven experimental treatments involving 692 subjects provide a systematic exploration of the existence and nature of motives for reciprocal behavior in two-person games. The experimental design supports discrimination between motivations of reciprocity and (non-reciprocal) altruism. The existence of reciprocal behavior is found to be dependent on the level of social distance but not on the level of monetary payoff. Alternative decision contexts such as framing the decision task as market exchange or eliciting strategy responses do not significantly change reciprocal behavior. However, the larger context in which a decision is made does have a significant effect on reciprocally-motivated behavior. These findings on payoff levels, social distance, decision context, and reciprocity have implications for a wide scope of research including both theoretical modeling and experimental design.

Keywords: game theory, reciprocity, altruism, experiments
JEL Classification: C70, C91, D63, D64

1. Introduction

The most widely-applied models in economics and game theory are based on the assumption of "self-regarding preferences," which are characterized by an exclusive motivation to maximize one's own material payoff. Models of self-regarding preferences capture behavior quite well in many types of situations, such as double auctions [Smith, 1982; Davis and Holt, 1993], one-sided auctions with independent private values [Cox and Oaxaca, 1996], procurement contracting [Cox, et al., 1996], and search [Cox and Oaxaca, 1989, 2000; Harrison and Morgan, 1990; Cason and Friedman, 2000]. But there is now a large body of literature that reports systematic inconsistencies with the implications of the self-regarding preferences model.[1] These replicable patterns of behavior are typically observed in experimental games involving salient decisions about the division of material payoffs among the experimental subjects. One explanation for the observed behavior that has received considerable attention is reciprocity. Our

study not only directly tests for reciprocity but also reveals how factors such as payoff levels, social distance, and decision context can affect reciprocal behavior.

We report experiments designed to yield insight into the nature of reciprocal motives. Only by observing decisions in a group of related experiments are we able to discriminate between behavior motivated by reciprocity and behavior motivated by non-reciprocal other-regarding preferences over outcomes. This is an often-overlooked design feature. Some treatments introduce the possibility of behavior motivated by positive reciprocity while other treatments introduce the possibility of negatively-reciprocal motivation. By "positive reciprocity," we mean a motivation to adopt a generous action that benefits someone else because that person's intentional behavior was perceived to be beneficial to oneself within the decision context of the experiment. Similarly, by "negative reciprocity" we mean a motivation to adopt a costly action that harms someone else because that person's intentional behavior was perceived to be harmful to oneself within the decision context of the experiment. Hence, in a given situation an action that would not otherwise be taken is considered reciprocal if it is undertaken in response to the action of another. Through additional treatments, we are able to study the robustness of reciprocity. The price elasticity of reciprocity is measured by doubling the monetary payoff level. By varying the social distance in the experimental protocol, specifically by using both single-blind and double-blind payoff procedures, we address the question whether social norms for reciprocal behavior are fully internalized or, alternatively, require external "enforcement" by the experimenters. Additionally we learn how changes in the decision context affect reciprocal behavior. Specifically, we compare sequential play of an extensive form game with strategy responses and sequential play with abstract and market-exchange framing. In total, the reported data from eleven experimental treatments involving almost 700 subjects constitutes a comprehensive and systematic investigation of reciprocity as a behavioral motivation.

Perhaps the most familiar experiment in the reciprocity literature is the ultimatum game. In this game, the first mover proposes a division of a fixed sum of money and the second mover

either accepts this proposal or vetoes it. In the event of a veto, both players get a money payoff of zero. The self-regarding-preferences model predicts extremely unequal payoffs for this game, as follows. Since the second mover is assumed to care only about maximizing her own money payoff, she should accept rather than veto any proposal that would give her a positive money payoff. Furthermore, the second mover is predicted to be indifferent between vetoing and accepting the proposal in which she receives a zero payoff. Knowing these preferences, and wanting only to maximize his own money payoff, the first mover is predicted to propose giving the second mover zero or the smallest feasible positive amount of money. However, observed behavior in the ultimatum game contrasts sharply with these predictions.

Under a wide variety of conditions, first movers in ultimatum games tend to propose relatively equal splits [Güth, Schmittberger, and Schwarze, 1982; Hoffman and Spitzer, 1985; Hoffman, McCabe, Shachat, and Smith, 1994; and Bornstein and Yaniv, 1998]. First movers may make generous proposals in ultimatum games because they have inequality-averse other-regarding preferences [Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000] or altruistic other-regarding preferences [Cox, Sadiraj, and Sadiraj, 2002b] over outcomes.[2] Alternatively, first movers may make generous proposals because they are afraid that second movers will veto lop-sided proposals. Second movers may veto such proposals because of inequality-averse preferences over outcomes or because of negative reciprocity. The implications for modeling behavior are different if the behavior is motivated by preferences over outcomes that are unconditional on perceived intentions than if it is motivated by negative reciprocity or fear of negative reciprocity. The latter requires the added complexity of state-dependent utility and a rule for determining the state in which the decision is made.[3] In order to discriminate among alternative motivations, we use a triadic experimental design that includes a mini-ultimatum game, which we call the Punishment mini-ultimatum game (Punishment-MUG), and two dictator control treatments.[4]

Additional insight into the nature of alternative motives is gained from a systematic comparison of our data with data from the different experimental design of Falk, Fehr, and Fischbacher [forthcoming], which also studies the Punishment-MUG. In order to more fully identify the causes of the behavioral differences, we run additional experimental treatments including framing the Punishment-MUG as market exchange and implementing the game with strategy method responses rather than sequential extensive form game responses.

Another game that is commonly studied in the reciprocity literature is the investment game of Berg, Dickhaut, and McCabe [1995] and a simplified version known as the Trust game. In contrast to the potentially negatively-reciprocal motives in the ultimatum game, play in the investment game could be motivated by second movers' positive reciprocity and first movers' trust that second movers will not defect because of their positive reciprocity and/or altruism. In the investment game, both the first and second movers are endowed with some amount of money. The second mover is instructed to keep his endowment while the first mover can keep her money or send any portion of it to the second mover. Any amount sent to the second mover is tripled by the experimenter. Subsequently, the second mover can keep the entire tripled amount or return any part of it to the first mover. The self-regarding preferences model predicts that second movers will keep all of the tripled amounts sent and, knowing this, first movers will send nothing. Observed behavior in the investment game contrasts sharply with these predictions of the self-regarding preferences model.

Berg, Dickhaut, and McCabe [1995] reported that, in their "no history" treatment, 28 out of 32 first movers sent more than the minimum possible positive amount of money ($1) to second movers. Out of the 28 second movers who received more than $3 (i.e., were sent more than the smallest possible positive amount), 11 shared the increase in total surplus by returning more than the first mover sent. First movers may make generous proposals in the investment game because they have altruistic other-regarding preferences over outcomes. Alternatively, first movers may make generous decisions because they trust that second movers will return enough money to

make the first mover payoff exceed the endowment. Second movers may return positive amounts because of altruistic other-regarding or inequality-averse preferences over outcomes. Alternatively, second movers may return money to first movers because of positive reciprocity.[5]

Again, the implications for modeling behavior are different if the behavior is motivated by preferences over outcomes that are unconditional on intentions than if it is motivated by positive reciprocity and trust in positive reciprocity. We use the Trust game, which is a commonly studied truncation of the investment game, to investigate the nature of positively-reciprocal motives. A triadic design, involving the Trust game and two dictator control games, is used to discriminate between positive reciprocity and non-reciprocal other-regarding preferences as motivations for decisions in the Trust game.[6] Other treatments are included that provide more insight into reciprocal motivations by varying the cost of generous decisions and varying the "social distance" implied by the experimental protocol. Specifically, we vary payoff levels by 100%, thus allowing an examination of the price elasticity of generous behavior, and we vary social distance by using both single-blind and double-blind payoff protocols to study whether social norms for positive reciprocity are internalized or require "enforcement" by interaction with the experimenter.

2. Discriminating Among Motives in the Punishment-MUG Game

In the Punishment-MUG shown in Figure 1, Mover 1 chooses either Take or Share. The top number in a terminal node payoff pair is Mover 1's dollar payoff, m_1 and the bottom number is Mover 2's dollar payoff, m_2. If Mover 1 chooses Take, then Mover 2 chooses between Tolerate and Punish. If Mover 1 chooses Share then Mover 2 chooses between Accept and Reject.

Mover 1 may choose Share rather than Take in the Punishment-MUG because he has altruistic or inequality-averse other-regarding preferences in which the outcome (m_1, m_2) =

($5,$5) is preferred to (m_1, m_2) = ($8,$2). Alternatively, Mover 1 may choose Share rather than Take because of fear that, if he chooses Take, Mover 2 will choose Punish. Similarly, Mover 2 may choose Punish rather than Tolerate because she has inequality-averse preferences for which the outcome, ($0,$0) is preferred to the outcome ($8,$2). Alternatively, Mover 2 may prefer the money payoff, ($8,$2) to ($0,$0) in the absence of a prior action by Mover 1 but choose Punish anyway because of negative reciprocity: a motivation to punish Mover 1 for attempting to obtain 80% of the total payoff for himself rather than making available the equal split. Thus if Mover 1 chooses Take, Mover 2 may be willing to incur the $2 cost of punishing this behavior even though she would prefer the outcome ($8,$2) to ($0,$0) in the absence of Mover 1's intentionally selfish choice.

The two punishment dictator control treatments shown in Figure 1 provide a way to discriminate among the alternative motivations. Here we provide an informal discussion of this central feature of the experimental design, leaving a formal analysis of the triadic design for the appendix. All decision choices are labeled with the name of the move in the Punishment-MUG. When the decision is in a control treatment, the name of the move is put in quotation marks. This convention allows decisions at corresponding nodes to be readily identified across treatments yet signifies that the named move is not characterized by the same motivations in the control treatment.

Punishment Control 1 is a dictator game in which Mover 1 chooses between "Take" and "Share" and the other player has no decision to make. In this game, Mover 1 can choose his preferred outcome pair without any fear of punishment by the other player. Thus the difference between first player decisions in the Punishment-MUG and decisions in the Punishment Control 1 game discriminates between Mover 1's choices of Share that are motivated by fear of negative reciprocity and those that are motivated by non-reciprocal other-regarding preferences.

Punishment Control 2 is a game in which the first move is determined by a random process. In this game, the "Take" branch and the "Share" branch have equal probability of being selected. Hence, if Mover 2 ends up choosing between ($8,$2) from "Tolerate" and ($0,$0) from "Punish," he has no reason to punish the other player for selfish play because Mover 1 has done nothing to put Mover 2 in this position. Thus the difference between second mover decisions in the Punishment-MUG and decisions in the Punishment Control 2 game discriminates between Mover 2's choices of ($0,$0) motivated by negative reciprocity and those motivated by non-reciprocal inequality-averse preferences. An alternative control for Mover 2's motives would be a simple dictator game in which Mover 2 decides between ($8,$2) and ($0,$0). However, as others have shown, behavior in mini-ultimatum games changes dramatically depending on the payoffs associated with unreached decision nodes [Falk, Fehr, and Fischbacher, forthcoming; Güth, Huck, and Müller, 2001]. Therefore, using this alternative control would confound the effects of mover 1's action and the elimination of a potential payoff pair.

3. Procedures in the Punishment-MUG Experiment

All participants were undergraduate students who received a $5 participation fee plus a payoff determined by playing one of the games. The experiments were computerized in order to minimize personal interaction between the researchers and the subjects and to maintain consistency across all sessions. The number of subjects in each of these laboratory sessions varied between twelve and twenty. Each subject participated once, in one game, and in only one session.

A double-blind protocol was used. Double-blind payoff procedures ensure that the only person who knows the decision of a specific individual is the individual herself. This was implemented by having subjects draw sealed envelopes from a box. Inside each envelope was a key labeled with an identification code. The subjects entered their identification codes into their

computers but did not enter any personal information. At the conclusion of the experiment, the subjects were escorted to a room containing similarly-coded, locked mailboxes. The researchers were not present in the room when subjects opened their mailboxes and retrieved plain white envelopes containing their money payoffs. The subjects were asked not to open their payoff envelopes until after leaving the building in order to maintain payoff privacy. This procedure is similar to that implemented by Berg, Dickhaut and McCabe [1995] and Cox [2002a,b].

The sequence of events within a session was carefully controlled to ensure both subject understanding and privacy. Upon arrival for the experiment, subjects waited in the sign-in room. Once all subjects were present, everyone received the participation fee and then entered the laboratory. Subjects were instructed to sit at any unoccupied station with a monitor displaying a light-gray screen. These stations were spaced evenly around the laboratory and separated by vacant stations. The vacant stations between subjects and the raised partitions on three sides of computer monitors provided significant privacy among subjects. After everyone was seated, the subjects began self-paced, computerized instructions. During this time, the experimenters answered subjects' questions. After all participants had completed the directions, a quiz was administered to ensure subject comprehension.[7] Anyone who answered the quiz incorrectly was given a private oral and graphical (game tree) explanation of the experiment by an experimenter. Following the oral explanations, the experimenters passed among the subjects with a box containing the sealed key envelopes. The subjects drew key envelopes from the box and were handed sheets of paper with written descriptions of the double-blind payoff procedures.[8] Subjects were told not to open their envelopes at this stage of the experiment. The description of the payoff procedures was read aloud to ensure that all participants knew that everyone was receiving the same information. At the conclusion of the reading, the subjects were told to open their envelopes after the experimenters had left the room.

One additional procedure was used in the experiment sessions with Punishment Control 2 only. In that treatment, a coin was flipped, in the presence of the subjects, once for each player 2

to determine nature's move. The procedure of flipping a coin has the advantages that subjects can verify the randomization process and the procedures are easy to comprehend. In all treatments, the experimenters left the laboratory and went to a monitor room before the subjects viewed the actual experiment game on their computer monitors. Each subject then made her decisions on the computer by clicking her mouse on the desired "tree branch" on the screen. The instructions and decision-making screens did not use the labels that appear on the game tree branches in Figure 1. Thus the subjects did not encounter evocative terms such as take, share, tolerate, punish, accept, and reject. Instead, they were presented with a game tree explained in neutral terms and the subjects were asked to use a mouse to click on the branch of their choice at decision nodes in the game tree. Games were referred to as decision trees while players were referred to as decision-makers.

As the amounts of money payoffs were determined by subjects' decisions, and appeared on the experimenters' monitor screen, money was inserted into envelopes by the experimenters in the monitor room. Once all subjects had made their decisions, the envelopes were put into the coded mailboxes in a room down a short hallway from the laboratory. Next, all computer screens were automatically blanked out to conceal subjects' decisions and subjects were asked to put their keys out of sight before they exited the laboratory to walk to the mailbox room. The experimenters did not enter the mailbox room while the subjects were there. After a total experiment time of approximately 45 minutes, subjects unlocked their mailboxes, retrieved their payoff envelopes, and exited.

4. Behavior in the Punishment-MUG Experiments

Figure 2 presents data from the experiments with the Punishment-MUG and its motivational controls. The results of two-sample proportions tests for behavioral differences between the Punishment-MUG and its controls are reported in Table 1. Note that 64% of the first movers choose Take in the Punishment-MUG and 70% choose "Take" in Punishment Control 1. Thus

more subjects choose the unequal division when the other player does not have the ability to punish than when she does, although the difference is not significant, as is evidenced by the z-statistic of -.5739. Therefore, we conclude that first movers' behavior in the Punishment-MUG is not characterized by fear of negative reciprocity. Furthermore, the first movers' revealed expectations in the Punishment-MUG are consistent with second movers' behavior because only 21% of second movers choose Punish. Thus choosing Take results in an average payoff of $6.36 for first movers whereas the choice of Share yields $4.77 on average.

While 21% of second movers in the Punishment-MUG did select Punish, this behavior cannot accurately be described as negative reciprocity. As explained above, the choice of Punish can only be interpreted as evidence of negative reciprocity if it is known that second movers prefer the outcome ($8,$2) to ($0,$0) when first movers have made no decision affecting the payoffs of second movers. Of the 13 Punishment Control 2 "second movers" who found themselves choosing between "Tolerate" and "Punish" because of the outcome of the coin flip, three (23%) chose "Punish." The difference between second mover behavior in the punishment and control games is in the wrong direction to support a conclusion that there is evidence of negative reciprocity in the Punishment-MUG, and the difference is insignificant (z-statistic = 0.196). Therefore we conclude that negative reciprocity does not explain behavior in this Punishment-MUG experiment. Second mover choices of ($0,$0) or ($5,$5) were not significantly different between the Punishment-MUG and the Punishment Control 2 treatments. More than 92% of the subjects that were required to make a decision selected outcome ($5,$5) over ($0,$0). The only possible surprise here is the 8% of the subjects that rejected ($5,$5). Interestingly, in a similar game Güth, Huck, and Müller [2001] also found occasional rejections of equal split offers.

5. Further Exploration of Motives in the Punishment-MUG

The conclusion that negative reciprocity and fear of negative reciprocity do not characterize behavior in the Punishment-MUG follows from comparison of behavior in the Punishment-MUG and the two dictator control treatments. Upon closer inspection, this surprising finding is due in part to our observing behavior that is fundamentally different from that reported by Falk, Fehr, and Fishbacker [forthcoming] for the Punishment-MUG. Table 2 compares behavior from the two studies. Based on the hypothesis tests presented in this table, the only decision at which players behaved similarly was the acceptance of ($5,$5).

At first our conclusion might appear to be based on fragile results. However, both studies have relatively large sample sizes for this type of research. This led us to ask what could explain the seemingly contradictory findings. As the answer should lie in the different experimental designs that were employed, we began to systematically investigate how the experiments differed. The Falk, et al. experimental design involved the strategy response mode; that is, they asked each subject to provide a complete contingency plan for each node at which the subject might have to make a decision. In contrast, our experimental design involved extensive form play of the game in which a second mover makes her decision after observing the first mover's decision. To determine whether this experimental design difference could account for the observed differences in behavior, we conducted another set of experiments with the Punishment-MUG in which thirty new subject pairs simultaneously entered complete strategy responses in their computers. The only changes from the protocol for the Punishment-MUG treatment detailed in the previous section were: (a) the requirement that a second mover enter responses for both mover 2 decision nodes in the Punishment-MUG (see Figure 1) prior to observing the first mover's decision; and (b) minimal wording changes in the directions that described how the outcome would be determined.

The results from our strategy method protocol for running the Punishment-MUG experiment were as follows. First, 18 out of the 30 first movers chose Take. Second, 25 out of

30 second movers selected Tolerate and all 30 of them chose to accept ($5,$5). As verified by the tests reported in Table 3, this behavior is identical to that observed in our sequential treatment. Therefore, we conclude that the sequential play and strategy method protocols do not elicit different patterns of behavior in our implementation of the Punishment-MUG. Data reported in Güth, Huck, and Müller [2001] also support the conclusion that sequential play and the strategy method elicit the same responses in mini-ultimatum games.[9] That study investigates an ultimatum game where the payoffs are double those of our Punishment-MUG and the Take payoff is slightly more lopsided in favor of the first mover.[10] Thus the available data suggest that the different results for the Punishment-MUG in our experiment and the Falk, et al. experiment cannot be attributed to the different message spaces used in the two experiments.

Another difference between the Falk, et al. design and ours was in the context in which the games were presented to the subjects. Instead of playing a single game, in Falk, et al. subjects simultaneously made decisions for the Punishment-MUG and three other games, only one of which would randomly be chosen to determine their payoff. The other three games were identical to the Punishment-MUG except that the ($5,$5) payoff was replaced with either ($10,$0), ($8,$2), or ($2,$8).[11] This difference in the context in which a game is played can significantly affect behavior. For example, Güth, et al. [2001] report significant differences in behavior between treatments in which individual games are played sequentially and treatments in which several games are played simultaneously using strategy responses.

The four games in the Falk, et al. experiment are presented in Figure 3. One way to characterize this difference between their design and ours is as a difference in how the Punishment-MUG is "framed." Perhaps framing effects account for the different patterns of behavior elicited by the two experiment protocols. In other words, behavior in the Punishment-MUG may not be robust to different ways of presenting the game to the subjects.

Hoffman, et al. [1994] introduced the market frame for eliciting decisions in an ultimatum game experiment and found that it led to more materially self-regarding behavior. We

implemented the market-exchange framing in the Punishment-MUG as follows. Player 1, the seller, sets a price for a good with zero marginal cost. Player 2, the buyer who values the good at $10, decides whether or not to make a purchase. The payoffs to the seller and the buyer are the price and $10 minus the price, respectively, if the good is sold and zero to both players otherwise. If framing effects are significant in the Punishment-MUG then one would expect even fewer rejections of ($8,$2) in favor of ($0,$0) when eliciting decisions with a market-frame protocol.

We conducted additional experiments with the Punishment-MUG using 30 new subject pairs, now framing the decisions as market exchange. Again, the same experimental procedures were employed except for the necessary changes in directions and computer interface.[12] Seventeen sellers set a price of $8, to which seven buyers responded by not purchasing. Of the thirteen sellers who set a price of $5, only one did not complete a sale. One immediate feature of this data is that a higher percentage of unequal offers were vetoed in the market frame than in the sequential treatment. This is in the opposite direction than that expected based upon the results of market-framed ultimatum game experiments reported by Hoffman, et al. [1994]. However, the market-frame treatment effect in the Punishment-MUG is not significant at the 90% confidence level, as reported in Table 3, which compares behavior at each node across the market-frame and game-tree protocols for the Punishment-MUG.

Taken as a whole, the data support the conclusion that negative reciprocity can be a significant motivational factor in the Punishment-MUG, but whether or not it is significant depends on the larger context in which the decisions are made. When subjects only make sequential decisions in the Punishment-MUG shown in Figure 1, their behavior is not characterized by negative reciprocity. Our exploratory treatments with strategy method responses and market framing, both concentrating exclusively on the Punishment-MUG, reveal that behavior in the Punishment-MUG is robust to these changes in experimental protocol. In contrast, when in the Falk, et al. experiment subjects simultaneously make strategy decisions in the Punishment-MUG and the three other games shown in Figure 3, their behavior is

characterized by significant negative reciprocity. A comparison of our data with data reported by Falk, et al. [forthcoming] reveals that significantly different behavioral patterns are elicited by the two different protocols for implementing the Punishment-MUG. Our data reveal that a first mover's choice of Take is not sufficient in itself to elicit significant negative reciprocity from a second mover in the Punishment-MUG. However, the Falk, et al. data reveal that when a second mover contemplates a first mover's choice of Take within the larger context of related games, significant negative reciprocity is elicited. The extent of this negatively-reciprocal behavior most likely depends on the specifics of the other games in the decision context. Güth, et al. [2001] also report the behavior of subjects making decisions in multiple related games played simultaneously. As opposed to Falk, et al. who compare the Punishment-MUG to games where the equal split is replaced by divisions at least as lopsided as (8,2), Güth, et al. change the equal split to only slightly favor one side or the other. Together, the data from those two studies indicate that the greater the asymmetry of the alternative feasible outcome, the less likely people are to choose Punish in response to Take.

6. Discriminating Among Motives in the Trust Game

In the Trust game presented in Figure 4, Mover 1 chooses between Exit and Engage. If Mover 1 chooses Exit then both players receive $5. If Mover 1 chooses Engage then Mover 2 chooses either Cooperate or Defect. If Mover 2 chooses Cooperate then the first mover receives $7.50 and the second mover receives $12.50. If Mover 2 chooses Defect then the first mover receives $0 and the second mover receives $20.

Mover 1 may choose Engage rather than Exit in the Trust game because she has altruistic preferences in which the outcomes ($7.50,$12.50) and ($0,$20) are *both* preferred to the outcome ($5,$5). Alternatively, Mover 1 may choose Engage because she trusts that Mover 2 will not defect, thus yielding the payoffs ($7.50,$12.50) that she prefers to ($5,$5). Mover 1 may choose Exit because she is afraid of defection by Mover 2, even though she prefers the payoffs

($7.50,$12.50) to the payoffs ($5,$5). Alternatively, Mover 1 may choose Exit because she has preferences with such strong inequality aversion that she prefers ($5,$5) to *both* ($7.50,$12.50) and ($0,$20).

Mover 2 may choose Cooperate rather than Defect in the Trust game because he has inequality-averse or altruistic other-regarding preferences in which the outcome, ($7.50,$12.50) is preferred to the outcome ($0,$20). Alternatively, Mover 2 may prefer the outcome ($0,$20) to the outcome ($7.50,$12.50) in the absence of a prior action by Mover 1 but choose Cooperate anyway because of a social norm for positive reciprocity. This can be explained as follows. Since Mover 2 observes the complete Trust game tree, she knows that Mover 1 chose between Exit, which would have assured Mover 1 of a payoff of $5, and Engage, which exposes Mover 1 to the risk of receiving a payoff of $0. If Mover 1 chooses Engage, Mover 2 may be willing to accept a payoff of $12.50 rather than getting $20 because of a social norm for positive reciprocity.

The trust dictator control treatments shown in Figure 4 provide a way to discriminate among alternative motivations. Again, we provide an informal discussion of this feature of the experimental design and leave formal analysis of the triadic design for the Trust game experiment to the appendix. All decision choices are labeled with the name of the move in the Trust game. When the decision is in a control treatment, the name of the move is put in quotation marks. As before, this convention allows decisions at corresponding nodes to be readily identified across treatments yet signifies that the name is not descriptive of motivation in the control treatments.

Trust Control 1A is a dictator game in which Mover 1 chooses between ($5,$5) and ($0,$20). It appears to be a reasonable judgment that any Mover 1 that chooses the monetary payoff pair ($0,$20) instead of ($5,$5) would also choose ($7.50,$12.50) instead of ($5,$5). However, if choices of ($0,$20) were to be observed in Trust Control 1A then it would be necessary to experiment with Trust Control 1B in order to discriminate between first-mover motives in the Trust game. The reason is that for a choice of Engage in the Trust game to be interpreted as a trusting action, it must be known that ($5,$5) is preferred to at least one of the

payoff pairs that is possible as a result of the player choosing Engage. In the event that all or nearly all subjects in Trust Control 1A choose ($5,$5), the difference between first player decisions in the Trust game and Trust Control 1A discriminates between Mover 1's choices of Engage that are motivated by trust in positive reciprocity and those that are motivated by non-reciprocal other-regarding preferences.

Trust Control 2 is a dictator game in which Mover 2 selects between "Cooperate," which results in the payoff pair ($7.50, $12.50), and "Defect," which yields Mover 2 a $20 payoff and results in a zero payoff for Mover 1. Comparing behavior differences between this allocation decision and the same allocation decision in the Trust game distinguishes behavior motivated by reciprocity from behavior motivated by non-reciprocal other-regarding preferences, since there is no decision by Mover 1 for which to reciprocate in Trust Control 2.

The experimental procedures used in the Trust game experiments were exactly the same as the procedures used in the sequential extensive form Punishment-MUG experiments except that there was no coin flip required for the second-mover control treatment.

7. Behavior in the Trust Game Experiments

The experiment results for the Trust game and its controls are reported in Figure 5. Several features of the data are readily apparent. First, only four out of the thirty subject pairs playing the Trust game reached the mutually beneficial ($7.50, $12.50) outcome, implying that this is not an environment in which people are very cooperative. A second prominent feature of the data is that a higher proportion of "second movers" chose "Cooperate" in the dictator control treatment than in the Trust game. Cooperate was observed for 24% of the Trust game second mover choices, whereas "Cooperate" was observed at a rate of 33% in Trust Control 2. This result is clearly inconsistent with reciprocity being the motivation for subjects' choice of Cooperate in the Trust game. If the decision to cooperate is the result of reciprocity, then in Trust Control 2 a lower frequency, not a higher frequency, of "Cooperate" should be observed. Table 4

contains the results of significance tests for equality of response rates among the observations in the Trust game and the two controls studied in the laboratory. With a two-sample proportions test, the z-statistic is 0.7062, hence the two rates are not significantly different from each other. We conclude that positive reciprocity does not explain behavior in this extensive form Trust game experiment.

The third main finding is that only two out of 30 Trust Control 1A "first movers" selected the ($0, $20) outcome over ($5, $5), whereas 17 out of 30 of the first movers in the Trust game chose Engage. Observed differences between Trust Control 1A and the complete Trust game provide evidence that, in the Trust game, first mover choices of Trust are, in fact, motivated by trust. The conclusion that trust motivates the behavior of first movers in the Trust game holds at all reasonable levels of significance, as the Z-statistic reported in Table 4 is 4.16. However, this trusting behavior is misplaced because second movers chose Defect at a high rate in the Trust game. In fact, given the observed player 2 behavior, the choice of Engage yields an expected monetary payoff of only $1.80 to first movers, whereas the choice of Exit yields a certain payoff of $5.

8. Further Exploration of Motives in the Trust Game

Our conclusion that positive reciprocity does not motivate behavior is due in part to the low frequency of Cooperate we observed in the Trust game. The 24% cooperation and 76% defection rates in our experiment are a reversal of the 75% cooperation and 25% defection rates reported by McCabe and Smith [2000].[13] Again, the question of why our data differ from another study arises. These experiments differed in two dimensions. First, the payoffs in their experiment were double those of our Trust game shown in Figure 4. Second, their experimental protocol had less "social distance" than ours. Their protocol had less social distance between the subjects and the experimenters and among the subjects, for the following reasons. They used a single-blind payoff procedure in which the subjects did not know the identity of their counterparts

but the experimenters knew which individuals made which decisions. Also, the subjects were paid in person, face to face, by the experimenters. Additionally, the McCabe and Smith experiment used exactly six subject pairs in the laboratory for each session, making it more likely that a subject would be matched with any other particular subject in the laboratory. In contrast, our laboratory sessions involved fourteen to twenty subjects. Each of these features of the McCabe and Smith protocol could have made their subjects more concerned than were our subjects about how others might judge their behavior.

Before investigating the effects of these protocol differences, we note that the high cooperation rate reported by McCabe and Smith [2000] was based on the responses of only twelve individuals. Therefore, we first attempt to replicate the results of that study, which coincidentally was conducted in the same laboratory as our experiments. Our replication consisted of four sessions with six subject pairs per session and used the same Novanet software as their study. Also, their single-blind procedure of calling subjects by name to come and individually receive payoffs directly from the experimenters was used. Table 5 compares the data from our original Trust game with data from McCabe and Smith [2000] and our replication of their experiment. The results support the conclusion that the findings of McCabe and Smith are not the result of a small sample size but, instead, are characteristic of behavior under their protocol. Therefore, the data from McCabe and Smith and our replication are combined in subsequent analysis.

Having replicated McCabe and Smith's results, we return our attention to the two design differences that might account for the behavioral shift, social distance and payoff level. To identify which design feature caused the behavioral change, we conducted an additional set of experiments with 27 new subject pairs. In these Trust game experiments, the payoffs were double those shown in Figure 4 and the level of social distance was kept high as more than twelve subjects participated in each session and the double-blind payoff procedure was again employed; thus these experiments differed in only one way from either of the two previously-discussed Trust

game experiments. In this double-blind, full-payoff experiment, 14 of the 27 first movers chose Engage. In response, 10 of the 14 second movers who had a decision to make chose Defect. Table 6 compares these Trust game data with data from the two social-distance, payoff-level combinations already reported. Behavior is statistically indistinguishable between the double-blind, half-payoff treatment (our original Trust game) and the double-blind, full-payoff treatment. But behavior is significantly different between the double-blind, full-payoff treatment and the single-blind, full-payoff treatment (the game of McCabe and Smith and our replication). Therefore, we conclude that it is the level of social distance, not the payoff level, that accounts for the behavioral differences between our original findings and the data reported by McCabe and Smith.

Because behavior is found to be contingent on social distance, a new question arises. Specifically, when the level of social distance is low, does positive reciprocity motivate behavior? So far none of our results are able to address this question. The reduction in social distance that leads more people to Cooperate in the Trust game may also lead more people to choose "Cooperate" in Trust Control 2. Because our objective is discrimination among behavioral motives, additional experiments were conducted with twenty-four new subject pairs playing Trust Control 2 for twice the payoffs shown in Figure 4, using the low-social-distance protocol of McCabe and Smith. The frequency with which these "second movers" selected "Cooperate" was identical to the rate reported earlier in Figure 5 for "second movers" in the original Trust Control 2, 33.33% in each treatment. Therefore, we conclude that behavior motivated by non-reciprocal other-regarding preferences remains unchanged with respect to altering the level of social distance. This finding also suggests that the double blind payoff procedure did not induce selfish behavior as a result of subjects questioning the reason for the high level of privacy and concluding that the experimenters wanted to observe such behavior. A comparison of "second mover" behavior in Trust Control 2, with low social distance and high payoffs, with second mover behavior in the low-social-distance, high-payoff version of the Trust game results in a z-

statistic of 2.43, which is significant at the 95% confidence level. Based on this test, we conclude that positive reciprocity does motivate behavior in the low social distance environment. Because there are no behavioral changes for first movers across the different Trust game protocols, and because altruism is shown not to be dependent on social distance and payoff level, we conclude that trust is present in the low social distance environment as well, and is a robust behavioral phenomenon.

From these six different experimental treatments involving the Trust game, we learn that positive reciprocity motivates behavior when the level of social distance is low but not when social distance is high. This implies that positive reciprocity is not an internalized norm that is triggered solely by the actions of others. Rather, positive reciprocity is a behavioral pattern that is dependent on the social context in which decisions are elicited. There is significant positive reciprocity observed in the Trust game when the experimenter can personally identify the subjects' actions and pays them their earnings, face to face. There is not significant positive reciprocity observed in the Trust game when the experiments are run with a double-blind protocol.[14]

9. Summary and Conclusions

We report results from experiments with 692 subjects and eleven treatments designed to explore conditions under which positive and negative reciprocity are and are not observed in two-person extensive form games. A triadic design, which compares decisions in situations with and without the potential for reciprocal motivation, is used to discriminate between behavior motivated by reciprocity and behavior motivated by non-reciprocal other-regarding preferences over outcomes. The treatments were designed to provide insight into the nature of reciprocal motives by varying the social distance incorporated into the experimental protocol, the framing of the decision task, and the monetary payoff level. Our data, together with data from two related studies, provide considerable new information about the intricacies of human behavior, in

particular the conditions under which reciprocal motives significantly affect behavior and the nature of those motives.

The Punishment game, a version of a mini-ultimatum game, was used to study negative reciprocity and fear of negative reciprocity. All of the treatments for the Punishment-MUG used a double-blind protocol in which subjects' decisions were anonymous both to other subjects and to the experimenters. Our initial treatments for the Punishment-MUG used sequential responses in which a second mover observed the decision of the first mover before making a decision. In this experiment, there were no significant differences between subjects' decisions in the Punishment-MUG and the two dictator control treatments. Thus, negative reciprocity was not observed in the behavior of second movers and fear of negative reciprocity was not observed in the behavior of first movers. Behavior in our initial Punishment-MUG experiment was significantly different than Punishment-MUG behavior in experiments reported by Falk, Fehr, and Fischbacher [forthcoming], who did observe significant negative reciprocity. The causes of these behavior differences were explored with several other experimental treatments.

The Falk, et al. experiment used strategy responses in which the "second mover" entered decisions for all possible nodes in a game without observing the first mover's decision. Our strategy response treatment explored whether this different message space produced different behavior in the Punishment-MUG. Because the data were not significantly different from our initial Punishment-MUG data, we concluded that use of strategy responses instead of sequential play of the game did not explain the differences in observed levels of negatively-reciprocal behavior. Our market exchange treatment explored whether framing the game as market exchange would produce different behavior in the Punishment-MUG. The data from the market-exchange treatment were also not significantly different from the data of our initial sequential treatment; hence the absence of negatively-reciprocal behavior in the Punishment-MUG was robust across market and non-market frames for sequential play.[15] In a different way from our triadic design, the Falk, et al. design also discriminates between behavior motivated by

reciprocity and behavior motivated by non-reciprocal preferences over outcomes. Based on our findings, it appears that the feature of their design that may account for significant negatively-reciprocal behavior is the simultaneous play of four games, as shown in Figure 3. Such play may make the choice of Take in the Punishment-MUG seem to be a more blatantly selfish action. Results reported by Güth, Huck, and Müller [2001] provide support for the conclusion that *simultaneous play* of several games with the strategy response mode elicits behavior that is significantly different than what is observed with either: (a) play of a single game with the strategy response mode; or (b) sequential (extensive form) play of a single game.

We used the Trust game, a truncated form of the investment game, to study positive reciprocity and trust in positive reciprocity. Our initial experiments with the Trust game used a double-blind protocol in which subjects' decisions were anonymous both to other subjects and to the experimenters. All of our experiments with the Trust game used sequential responses in which a second mover observed the decision of the first mover before making a decision. In the initial experiment, there was a significant difference between first movers' play in the Trust game and play by "first movers" in the relevant dictator control treatment. Thus the first movers' behavior was characterized by significant trust in positive reciprocity. But there was not a significant difference between second movers' play in the Trust game and play by "second movers" in the control treatment. Thus second mover behavior was not characterized by significant positive reciprocity. Again, behavior in our initial experiments differed significantly from work previously reported, in this case by McCabe and Smith [2000]. The causes of these behavioral differences were also explored with other experimental treatments.

Our initial Trust game experiments used money payoffs that were one-half the payoffs used by McCabe and Smith. Also, our experiments differed in the level of social distance; we used a double-blind payoff procedure whereas they used a single-blind protocol, and our experiment sessions involved a larger number of subjects in the laboratory. In order to investigate the effects of payoff level and social distance, we first replicated the results reported

by McCabe and Smith. Next, we ran another experiment with the Trust game that differed from our initial one in that payoffs were twice as large. This experiment differed from McCabe and Smith's experiment in that the social distance was increased. Data from this treatment were not significantly different from data in our initial Trust game experiment but the data were significantly different from the data of McCabe and Smith combined with data from our replication of their experiment. Therefore we conclude that the payoff level does not account for the differences in observed behavior but rather the level of social distance drives the results. This conclusion is consistent with the earlier finding by Hoffman, et al. [1994] that single-blind and double-blind payoff protocols can elicit significantly different behavior in some contexts.

Having established that cooperation was dependent on social distance, we next asked whether the low social distance protocol elicited positively reciprocal behavior or, alternatively, if this protocol would produce similar shifts in behavior in both the Trust game and the second-mover control treatment. In order to answer this question, we ran the second-mover control treatment with the doubled payoff level and the low-social-distance protocol. Data from this implementation of the control treatment were not significantly different from data for the initial implementation of the second-mover control game, but the data were significantly different from the low-social-distance, high-payoff Trust game data; that is, we observed significant positive reciprocity in this environment. Thus we concluded that the social norm for positive reciprocity is not sufficiently internalized to produce positively-reciprocal behavior in the Trust game when a high-social-distance protocol is used. Instead, social "enforcement" by the experimenters' personal knowledge of subjects' behavior is necessary to elicit positively-reciprocal behavior in the Trust game.

A different approach to varying the level of social distance as an experimental treatment was implemented by Charness, Haruvy, and Sonsino [2001]. They ran strategy-method response-mode experiments with the wallet game in which the two treatments were: (a) laboratory experiments with double-blind payoffs; and (b) Internet experiments.[16] In the laboratory

treatment, the subjects knew they were all from the same social group, "students," and were assembled in one room. In contrast, in the Internet treatment the social group composition of participants was unknown to the subjects and there was greater anonymity because of the physical dispersion of the subjects. Charness, et al. report a significant difference between data from the two treatments.

There is now a large literature on experiments that produce data that are inconsistent with the traditional self-regarding preferences model in which an agent's only motivation is to maximize his own material payoffs. The experiments that produce most of these data involve games with obvious fairness considerations in the subjects' decision tasks. Some of the reported experimental designs, including ours, can discriminate among some alternative causes of the deviations from self-regarding behavior. Thus, our triadic design discriminates between behavior motivated by trust or reciprocity and behavior motivated by other-regarding preferences characterized by altruism [Cox, Sadiraj, and Sadiraj, 2002b], maxi-min considerations [Charness and Rabin, forthcoming], or inequality aversion [Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000]. Such discrimination provides insight into the nature of the reciprocal motives that must be incorporated into game theory in order to improve the theory's empirical validity. Furthermore our design, together with others in the literature, yield subtle insights into the nature of reciprocal motives.

Together, our experimental design and those reported by Falk, et al. [forthcoming] and Güth, et al. [2001] provide support for the conclusion that reciprocal behavior can differ significantly between experimental designs involving sequential (extensive form) play of a single game and designs involving simultaneous play of several games with strategy responses.

Our experimental design and the design in Hoffman, et al. [1994] reveal that the difference in social distance between single-blind and double-blind payoff protocols in laboratory experiments can significantly affect behavior. The Charness, et al. [2001] experimental design

reveals that the difference in social distance between double-blind laboratory experiments and Internet experiments can also significantly affect behavior.

An extensive literature establishes that reciprocal motivations significantly affect behavior in controlled experiments. The next stage of the research involves exploration of the nature of the reciprocal motives. The present paper is a contribution to that exploration.

Appendix: Discriminating Among Motives with the Triadic Design

1. The Punishment-MUG and its Controls

The Punishment-MUG is a simplified version of the familiar ultimatum game. In this game Player 1 can propose to either divide \$10 evenly or to keep \$8 for themselves. Player 2 can respond to the proposed split by either agreeing to it or by vetoing it in favor of a \$0 payoff to each player.

The Punishment-MUG

In the Punishment-MUG, shown in Figure 1, Mover 1 chooses either Take or Share. If Mover 1 chooses Take, then Mover 2 chooses between Tolerate and Punish. If Mover 1 chooses Share then Mover 2 chooses between Accept and Reject.

At the time Mover 1 contemplates his choice, he may not know what choice Mover 2 will subsequently make if the first mover's choice is Take. Let $\pi_P \in [0,1]$ be Mover 1's subjective probability that Mover 2 will choose Punish. Normalize Mover 1's utility function so that $u^1(0,0) = 0$. Then Mover 1's expected utility from choosing Take is

(1) $\quad U_P^1(T) = (1-\pi_P)u^1(8,2)$.

Let $\pi_R \in [0,1]$ be Mover 1's subjective probability that Mover 2 will choose reject if Mover 1 chooses Share; then Mover 1's expected utility from choosing Share is

(2) $\quad U_P^1(S) = (1-\pi_R)u^1(5,5)$.

The choice of Take by Mover 1 may trigger a social norm for negative reciprocity in Mover 2. To allow for that possibility, we let the utility of Mover 2 depend on a state variable, $\lambda(T)$ in the event that Mover 1 chooses take. Thus the utility to Mover 2 from choosing Punish is

(3) $\quad U_P^2(P) = u_{\lambda(T)}^2(0,0)$.

The payoff to Mover 2 from choosing Tolerate is

(4) $\quad U_P^2(T) = u_{\lambda(T)}^2(2,8)$.

Punishment Control 1: Dictator Control Treatment for Fear

The dictator control treatment for fear of punishment is also illustrated in Figure 1. Punishment Control 1 is a dictator game in which Mover 1 chooses between "Take" and "Share." The utility to Mover 1 will thus be either $u^1(8,2)$ or $u^1(5,5)$.

Punishment Control 2: Dictator Control Treatment for Negative Reciprocity

Punishment Control 2, illustrated in Figure 1, is a game in which nature randomly selects which of two dictator games Mover 2 will play. Mover 2 has a 0.5 probability of choosing between "Tolerate" and "Punish" and a 0.5 probability of choosing between "Accept" and "Reject." Since Mover 1 has no decision to make in Punishment Control 2, the second mover's utility will not depend on a social norm for negative reciprocity. Mover 2's utility from choosing "Tolerate" is

(5) $U^2_{PC2}(T) = u^2(2,8)$.

The utility to Mover 2 from choosing "Punish" is

(6) $U^2_{PC2}(P) = u^2(0,0)$.

Testing for the Presence of Fear

In order to conclude that a first mover has exhibited fear of punishment, the researcher must have knowledge that Mover 1 has revealed his belief that the probability that Mover 2 will Punish, π_P, is larger than the probability that Mover 2 will Reject, π_R. This is implied by the first mover's choice of Share in the Punishment-MUG and "Take" in Punishment Control 1, as shown by the following. The choice of Share in Punishment-MUG implies

(7) $(1-\pi_R)u^1(5,5) \geq (1-\pi_P)u^1(8,2)$.

The choice of "Take" in Punishment Control 1 implies

(8) $u^1(5,5) \leq u^1(8,2)$.

Statements (7) and (8), together with the assumed absence of indifference between (5,5) and (8,2), imply

(9) $\pi_P > \pi_R$.

Testing for the Presence of Negative Reciprocity

In order to conclude that a second mover has demonstrated negative reciprocity, the researcher must know that the second mover has borne a cost to inflict a loss on the first mover. This can be manifested by the second mover making a choice in the Punishment-MUG that is known to be disprefered to the alternative choice in the absence of a social norm for negative reciprocity.

The choice of Punish in the Punishment-MUG reveals that

(10) $u^2_{\lambda(T)}(0,0) \geq u^2_{\lambda(T)}(2,8)$.

The choice of "Tolerate" in Punishment Control 2 reveals that

(11) $u^2(2,8) \geq u^2(0,0)$.

Statements (10) and (11) are consistent only because of the presence in (11) of the state variable for negative reciprocity.

2. The Trust Game and its Controls

The Trust game is a simplified version of the common investment game. In this game Player 1 can decide to end the game with both players keeping their $5 endowments or Player 1 can pass their entire $5 endowment to Player 2. If the money is passed to Player 2 it is tripled and Player 2 can decide to keep the additional $15 for themselves or can return $7.50 to Player 1.

The Trust Game

In the Trust game, graphically presented in Figure 4, Mover 1 chooses either Exit or Engage. If Mover 1 chooses Exit then both players receive $5, with utility payoffs,

(12) $U^i_T = u^i(5,5)$, $i = 1,2$.

If Mover 1 chooses Engage then Mover 2 chooses either Cooperate or Defect. If Mover 2 chooses Cooperate then the first mover receives \$7.50 and the second mover receives \$12.50. If Mover 2 chooses Defect then the first mover receives \$0 and the second mover receives \$20.

At the time Mover 1 contemplates her choice, she may not know what choice Mover 2 would subsequently make if the first mover's choice is Engage. Let $\pi_C \in [0,1]$ be Mover 1's subjective probability that Mover 2 will choose Cooperate; then Mover 1's expected utility from choosing Engage is

(13) $U_T^1(E) = \pi_C u^1(7.50,12.50) + (1-\pi_C)u^1(0,20)$.

The choice of Engage by Mover 1 may trigger a social norm for positive reciprocity in Mover 2. To allow for that possibility, we let the utility of Mover 2 depend on a state variable, $\lambda(E)$ in the event that Mover 1 chooses Engage. Thus the payoff to Mover 2 from choosing Cooperate is

(14) $U_T^2(C) = u_{\lambda(E)}^2(12.50,7.50)$.

The payoff to Mover 2 from choosing Defect is

(15) $U_T^2(D) = u_{\lambda(E)}^2(20,0)$.

Trust Control 1A & 1B: Dictator Control Treatments for Trust

The trust dictator control games are also illustrated in Figure 4. Trust Control 1A is a dictator game in which Mover 1 chooses between "Exit," with payoff $u^1(5,5)$, and "Engage," with payoff $u^1(0,20)$. It appears to be a reasonable judgment that any Mover 1 that chooses the monetary payoff pair, (0, 20) instead of (5, 5) would also choose
(7.50, 12.50) instead of (5, 5). However, if the choice of (0, 20) is observed in Trust Control 1A then it would be necessary to experiment with Trust Control 1B in order to obtain all of the relevant information about Mover 1 preferences.

Trust Control 2: Dictator Control Treatment for Positive Reciprocity

Trust Control 2, shown in Figure 4, is a dictator game in which Mover 2 chooses between "Cooperate" and "Defect." Since Mover 1 has no decision to make in Trust Control 2, the second mover's utility will not depend on a social norm for positive reciprocity. Mover 2's utility from choosing Cooperate is

(16) $U^2_{TC2}(C) = u^2(12.50,7.50)$.

The utility to Mover 2 from choosing Defect is

(17) $U^2_{TC2}(D) = u^2(20,0)$.

Testing for the Presence of Trust

In order to conclude that a first mover has demonstrated trust, the researcher must have knowledge that player 1 has borne a risk of loss from her choice in the Trust game. A first mover that has chosen Engage in the Trust game will have demonstrated trust if it is known that

(18) $u^1(5,5) > \min[u^1(7.50,12.50), u^1(0,20)]$

The selection of Trust Control 1A as the control treatment for trust is based on the assumption that for first movers it will be true that $u^1(7.50,12.50) > u^1(0,20)$. The choice of (y^1, y^2) = (5, 5) over (y^1, y^2) = (0, 20) in Trust Control 1A reveals that

(19) $u^1(5,5) \geq u^1(0,20)$.

Given that it is unlikely that a subject is indifferent between (5, 5) and (0, 20), we conclude that a subject has exhibited trust if he chooses Engage in the Trust game and "Exit" in Trust Control 1A.

Testing for the Presence of Positive Reciprocity

In order to conclude that a second mover has demonstrated positive reciprocity, the researcher must know that the second mover has incurred a cost to repay a social debt to the first mover. This can be manifested by the second mover making a choice in the Trust game that is

known to be dispreferred to the alternative choice in the absence of a social norm for positive reciprocity. The choice of (y^1, y^2) = (7.50, 12.50) in the Trust game reveals that

(20) $u^2_{\lambda(E)}(12.50,7.50) \geq u^2_{\lambda(E)}(20,0)$.

The choice of (y^1, y^2) = (0,20) in Trust Control 2 reveals

(21) $u^2(20,0) \geq u^2(12.50,7.50)$.

Given that it is unlikely that a subject is indifferent between (y^1, y^2) = (7.50, 12.50) and (y^1, y^2) = (0,20), we conclude that a subject has exhibited positive reciprocity if she chooses Cooperate in the Trust game and "Defect" in Trust Control 2.

References

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

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

Bornstein, Gary and Ilan Yaniv, "Individual and Group Behavior in the "Ultimatum Game: Are Groups More 'Rational' Players?" *Experimental Economics*, I(1998), 101-08.

Cason, Timothy and Daniel Friedman, "Buyer Search and Price Dispersion: A Laboratory Study," Discussion paper, University of California at Santa Cruz, July 2001.

Charness, Gary, Ernan Haruvy, and Doron Sonsino, "Social Distance and Reciprocity: The Internet vs. the Laboratory," Discussion paper, University of California at Santa Barbara, 2001.

_____ and Matthew Rabin, "Social Preferences: Some Simple Tests and a New Model," *Quarterly Journal of Economics*, forthcoming.

Cox, James C., "Trust, Reciprocity, and Other-Regarding Preferences: Groups vs. Individuals and Males vs. Females," in *Advances in Experimental Business Research*, Rami Zwick and Amnon Rapoport, eds. (Boston, MA: Kluwer Academic Publishers, 2002a).

_____, "How to Identify Trust and Reciprocity," University of Arizona discussion paper, May 2002b.

_____, R. Mark Isaac, Paula-Ann Cech, and David Conn, "Moral Hazard and Adverse Selection in Procurement Contracting," *Games and Economic Behavior*, XVII(1996), 147-76.

_____ and Ronald L. Oaxaca, "Laboratory Experiments with a Finite Horizon Job Search Model," *Journal of Risk and Uncertainty*, II(1989), 301-29.

_____ and _____, "Is Bidding Behavior Consistent with Bidding Theory for Private Value Auctions?" *Research in Experimental Economics* VI, R. Mark Isaac ed. (Greenwich, CT: JAI Press, 1996).

_____ and _____ , "Good News and Bad News: Search from Unknown Wage Offer Distributions," *Experimental Economics*, II(2000), 197-225.

_____, Klarita Sadiraj, and Vjollca Sadiraj, "Trust, Fear, Reciprocity, and Altruism," University of Arizona working paper, May 2002a.

_____, _____, and _____, "A Theory of Competition and Fairness for Egocentric Altruists," University of Arizona working paper, June 2002b.

Davis, Douglas and Charles Holt, *Experimental Economics*, (Princeton, NJ: Princeton University Press, 1993).

Deck, Cary A., "A Test of Behavioral and Game Theoretic Models of Play in Exchange and Insurance Environments," *American Economic Review*, XCI(2001), 1546-55.

Falk, Armin, Ernst Fehr, and Urs Fischbacher, "On the Nature of Fair Behavior," *Economic Inquiry*, forthcoming.

Fehr, Ernst and Simon Gächter, "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives*, XIV(2000), 159- 81.

_____ and Klaus M. Schmidt, "A Theory of Fairness, Competition and Cooperation," *Quarterly Journal of Economics*, CXIV(1999), 817-68.

Güth, Werner, Steffen Huck and Wieland Müller, "The Relevance of Equal Splits in Ultimatum Games," *Games and Economic Behavior*, XXXVII(2001), 161-9.

_____, Rolf Schmittberger and Bernd Schwarze, "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization*, III(1982), 367-88.

Harrison, Glenn W. and Peter Morgan, "Search Intensity in Experiments," *Economic Journal*, C(1990), pp. 478-86.

Hoffman, Elizabeth, Kevin A. McCabe, Keith Shachat, and Vernon L. Smith, "Preferences, Property Rights, and Anonymity in Bargaining Games," *Games and Economic Behavior*, VII(1994), 346-80.

_____ and Matthew L. Spitzer, " Entitlement, Rights and Fairness: An Experimental Examination of Subjects' Concepts of Distributive Justice," *Journal of Legal Studies*, XIV(1985), 259-97.

McCabe, Kevin A. and Vernon L. Smith, "A Comparison of Naïve and Sophisticated Subject Behavior with Game Theoretic Predictions," *Proceedings of the National Academy of Sciences*, XCVII(2000), 3777-81.

McKelvey, Richard D. and Thomas R. Palfrey, "An Experimental Study of the Centipede Game," *Econometrica*, LX(1992), 803-36.

_____ and _____, "Quantal Response Equilibria for Extensive Form Games," *Experimental Economics*, I(1998), 9-41.

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

Smith, Vernon L., "Microeconomic Systems as an Experimental Science*," American Economic Review*, LXXII(1982), 923-55.

Endnotes

* We are grateful for research support from the Decision Risk and Management Science Program of the National Science Foundation (grant number SES9818561). Helpful comments and suggestions were provided by Gary Charness and Rachel Croson.

[1] Some of this literature is reviewed in Fehr and Gächter [2000].

[2] A person with inequality averse preferences receives positive marginal utility from the income of the lowest payoff agent and negative marginal utility from the income of the highest payoff agent. In contrast, a person with altruistic other regarding preferences receives positive marginal utility from the income of every agent, and these preferences are egocentric if the person prefers self-favoring unequal payoffs to other-favoring unequal payoffs of the same amounts.

[3] The appendix contains a model that represents the different effects on subjects' motivations, of the three treatments in a triadic design, with state-dependent utility functions.

[4] A triadic design decomposes motives in a two-player game by comparing behavior across three games: the original two-player game and two single-mover games in which the sole mover chooses among the outcome choices of the original game.

[5] Some other explanations for this behavior have been developed, including the game-theoretic implications of stochastic decision-making in the form of quantal response [McKelvy and Palfrey, 1992, 1998] and a specific model that incorporates intentions [Rabin, 1993]. Deck [2001] shows that these models of behavior as well as the aforementioned models of Fehr and Schmidt [1999] and Bolton and Ockenfels [2000] cannot fully explain subject decisions in several games, including a modified version of the Trust game used in the present paper.

[6] Cox [2002a,b] uses a triadic design to discriminate among motives in the investment game.

[7] The control experiments for negative reciprocity did not include a quiz. These experiments were completed before the quiz was designed. However, the full Punishment-MUG was

conducted both with and without the quiz and no significant difference between the two data sets was found. Therefore, re-running the control experiments was considered unnecessary.

[8] The subject instructions, quiz, and description of the double-blind payoff procedures are available online at comp.uark.edu/~cdeck/expinfo.htm.

[9] When the decision context involves subjects making decisions in a single game, the null hypothesis that the proportion of subjects selecting an action is the same across elicitation methods cannot be rejected at the 10% significance level in favor of the two sided alternative for any of the three decision nodes of the "Equal" game, which is similar to our Punishment-MUG.

[10] Our strategy method data are statistically indistinguishable from the Güth, et al. strategy method data, but we find less frequent rejection of the lopsided offer in the sequential game. However, this discrepancy could be due to the relatively small sample size reported in Güth, et al. where six of ten subjects choose to Punish.

[11] Note that the design of the Falk et al. experiments does allow one to conclude that negative reciprocity is observed if (8,2) is rejected in the 5/5 game but is accepted in the 8/2 game shown in Figure 3. This is the pattern of responses in their data.

[12] The directions for the market frame experiments closely followed Hoffman et al. [1994]. The displays of the computer screens were similar to our other treatments in layout, color, and format.

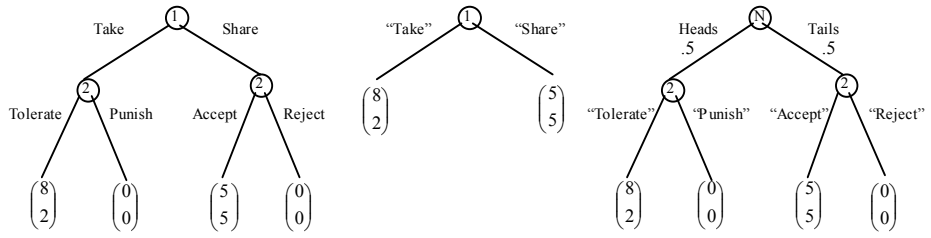
[13] The experimental design in McCabe and Smith [2000] cannot identify behavior as being motivated by reciprocity because no control experiments with the reciprocal motivations eliminated were included. Our paper addresses the existence of reciprocity in the environment of McCabe and Smith [2000] after first identifying the causes for the behavioral differences in the Trust game.

[14] This contrasts with some other games. Double-blind protocols have been found to elicit positively-reciprocal behavior in triadic-design experiments with the investment game [Cox, 2002a,b] and the moonlighting game [Cox, Sadiraj, and Saditraj, 2002a]. One notable difference between those games and the Trust game is that they have many more possible responses that the subjects can make. A denser message space may allow decision-makers to undertake less costly reciprocal behavior.

[15] Experiments with a triadic design involving sequential play of the moonlighting game and two control treatments [Cox, Sadiraj, and Sadiraj, 2002a] also produced data in which negatively-reciprocal behavior was insignificant.

[16] The wallet game is structured as follows: Mover 1 can either keep $x leaving Mover 2 with $0 or allow Mover 2 to decide how to allocate $y $\geq$ $x between the two movers. The game models the situation in which one person finds a missing wallet containing cash. The finder can keep the wallet or return it to the owner who values both the cash and the other contents of the lost wallet. Upon the return of the wallet the owner has the option to pay a reward to the finder.

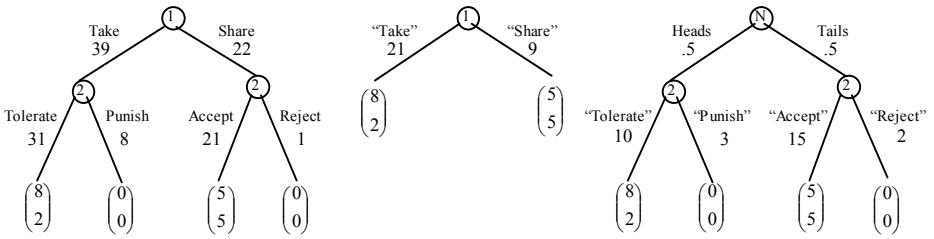
Figure 1. The Punishment-MUG Game and Dictator Controls



Punishment-MUG | Punishment Control 1 | Punishment Control 2

N denotes a move by nature. The numbers beside nature's decision branches indicate the probability of the move being selected by a coin fl ip.

Figure 2. Behavior in Punishment-MUG and Dictator Controls



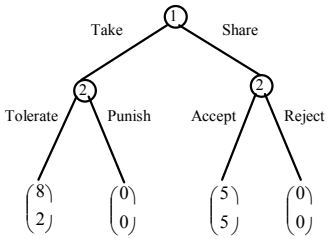
Punishment-MUG

Punishment Control 1

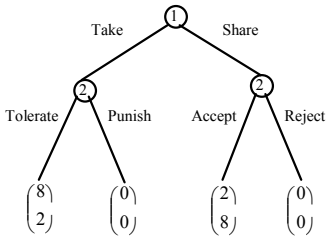
Punishment Control 2

Numbers beside each move indicate the number of subjects selecti ng that move. N denotes a move by nature. The numbers beside nature's decisio n branches indicate the probability of the move being selected by a coin fl ip.

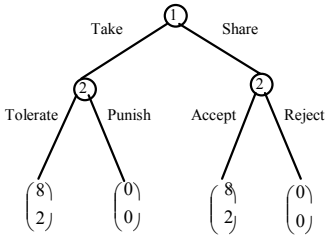
Figure 3. The Experiments of Falk, Fehr, and Fischbacher



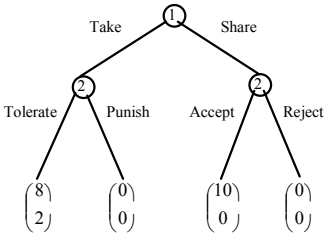
5/5 – game (Punishment)



2/8 – game



8/2 – game



10/0 – game

Figure 4. The Trust Game and Dictator Controls

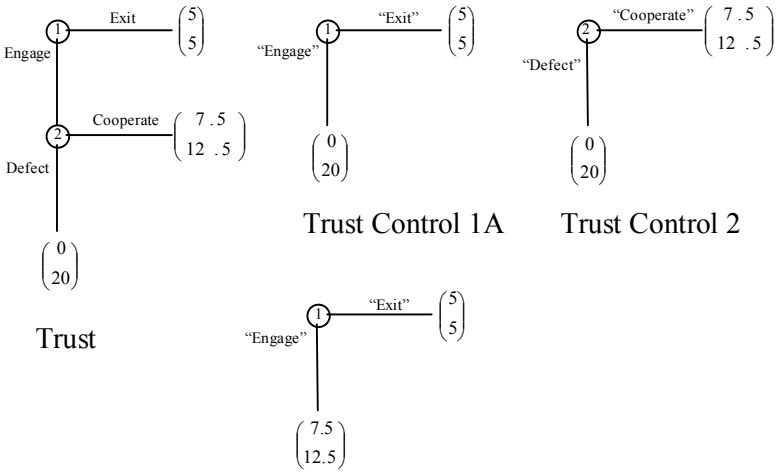
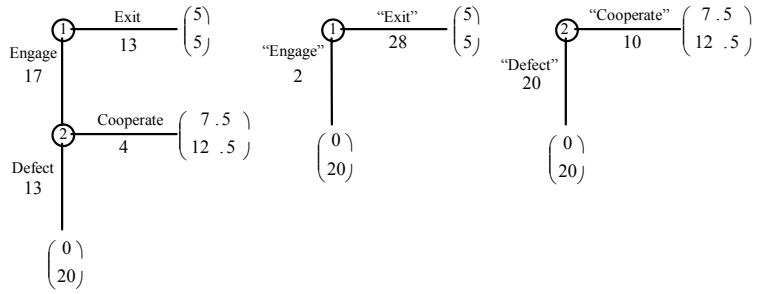


Figure 5. Behavior in Trust and Dictator Controls



Trust Trust Control 1A Trust Control 2

Numbers beside each move indicate the number of subjects selecting that move.

Table 1. Motive Discrimination for Behavior in the Punishment-MUG

Decision	Take vs "Take"	Punish vs "Punish"	Accept vs "Accept"
Proportion Making Decision in Punishment-MUG	0.64	0.21	0.95
Number of Observations in Punishment-MUG	61	39	22
Proportion Making "Decision" in Punishment Control Game	0.70	0.23	0.88
Number of Observations in Punishment Control Game	30	13	17
Estimate of Common Proportion	0.66	0.21	0.92
z-statistic	-0.574	-0.196	0.839

The z-statistic is for a two-sample proportion test where the null hypothesis is Ho: the proportion of subjects making the same decision is equal in the two data sets.

Table 2. Comparison of Punishment-MUG Results with Falk, et al. Data

Decision	Take	Punish	Accept
Proportion Reported Making Decision in Falk, et al.	0.31	0.44	1.00
Number of Observations in Falk et al.	45	45	45
Proportion Making Decision in Punishment-MUG in this study	0.64	0.21	0.95
Number of Observations in this study	61	39	22
Estimate of Common Proportion	0.50	0.33	0.99
z-statistic	-3.341	2.321	-1.441

The z-statistic is for a two-sample proportion test where the null hypothesis is Ho: the proportion of subjects making the same decision is equal in the two data sets.

Table 3. Comparison of Behavior in the Punishment-MUG
Between Decision Making Contexts

Decision	Take		Punish		Accept	
Alternative Context	Strategy Method	Market Frame	Strategy Method	Market Frame	Strategy Method	Market Frame
Proportion Making Decision in Alternative Context Punishment-MUG	0.6	0.57	0.17	0.41	1.00	0.92
Number of Observations in Alternative Context Punishment-MUG	30	30	30	17	30	13
Proportion Making Decision in (Sequential) Punishment-MUG	0.64	0.64	0.21	0.21	0.95	0.95
Number of Observations in (Sequential) Punishment-MUG	61	61	39	39	22	22
Estimate of Common Proportion	0.63	0.62	0.19	0.27	0.98	0.94
z-statistic	-0.365	-0.670	-0.405	1.606	1.179	-0.388

The z-statistic is for a two-sample proportion test where the null hypothesis is Ho: the proportion of subjects making the same decision is equal in the two data sets.

Table 4. Motive Discrimination for Behavior in the Trust Game

Decision	Engage vs "Engage"	Cooperate vs "Cooperate"
Proportion Making Decision in Trust Game	0.57	0.24
Number of Observations in Trust	30	30
Proportion Making "Decision" in Trust Control Game	0.07	0.33
Number of Observations in Trust Control	30	30
Estimate of Common Proportion	0.32	0.30
z-statistic	4.163	-0.706

The z-statistic is for a two-sample proportion test where the null hypothesis is Ho: the proportion of subjects making the same decision is equal in the two data sets.

Table 5. Comparison of Trust Game Results with McCabe and Smith [2000] Data

Game	Trust		McCabe and Smith Data		Replication of McCabe and Smith	
Decision	Observed Frequency	Sample Size	Observed Frequency	Sample Size	Observed Frequency	Sample Size
Engage	17	30	12	24	13	24
Cooperate	4	17	9	12	8	13

Game Comparison	Decision	Estimate of Common Proportion	z-statistic
Trust vs. McCabe and Smith Data	Engage	0.54	0.488
	Cooperate	0.49	-2.745
Trust vs. Replication of McCabe and Smith	Engage	0.52	0.792
	Cooperate	0.40	-2.106
Replication of McCabe and Smith vs. McCabe and Smith Data	Engage	0.48	0.289
	Cooperate	0.68	-0.721

The z-statistic is for a two-sample proportion test where the null hypothesis is Ho: the proportion of subjects making the same decision is equal in the two data sets.

Table 6. Effects of Social Distance and Payoff Level on Behavior in the Trust Game with High Social Distance Full Payoff as the Baseline

Decision	Engage		Cooperate	
Alternative Environment	High Social Distance Half Payoff	Low Social Distance Full Payoff	High Social Distance Half Payoff	Low Social Distance Full Payoff
Proportion Making Decision in Alternative Environment	0.57	0.52	0.24	0.68
Number of Observations in Alternative Environment	30	48	17	25
Proportion Making Decision in Baseline Environment	0.52	0.52	0.29	0.29
Number of Observations in Baseline Environment	27	27	14	14
Estimate of Common Proportion	0.54	0.52	0.26	0.54
z-statistic	0.364	0.019	-0.319	2.369

The z-statistic is for a two-sample proportion test where the null hypothesis is Ho: the proportion of subjects making the same decision is equal in the two data sets. High Social Distance Half Payoff is the environment used initially for the Trust game in this study. The Low Social Distance Full Payoff is the environment of McCabe and Smith [2000] and our replication of that study.