On the Nature of Reciprocal Motives

James C. Cox
and
Cary A. Deck*

Forthcoming in *Economic Inquiry*.

JEL C70, C91, D63, D64

On the Nature of Reciprocal Motives

Data from 692 subjects in 11 experimental treatments provide a systematic exploration of the existence and nature of reciprocal behavior in two-person games. The experimental design discriminates between motivations of reciprocity and (non-reciprocal) other-regarding preferences. The existence of positive reciprocity is found to be dependent on the level of social distance but not the level of monetary payoff. The larger context in which a decision is made is found to have a significant effect on negative reciprocity. These findings on payoff levels, social distance, decision context, and reciprocity have implications for both theoretical modeling and experimental design.

James C. Cox

University of Arizona

Cary A. Deck

University of Arkansas

I. 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 concern about one's own material payoff. Models of self-regarding preferences capture behavior quite well in many contexts, including double auctions as in Smith (1982) and Davis and Holt (1993), one-sided auctions with independent private values as in Cox and Oaxaca (1996), procurement contracting as in Cox, et al. (1996), and search as in Cox and Oaxaca (1989; 2000), Harrison and Morgan (1990) and 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 observed in experimental games involving decisions about the division of material payoffs among the participants. One explanation for the observed behavior that has received considerable attention is reciprocity.

We report experiments designed to yield insight into the nature and robustness of reciprocal motives. By observing decisions in a group of related experiments we are able to discriminate between behavior motivated by reciprocity and behavior motivated by non-reciprocal other-regarding preferences over outcomes. 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.

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, with the first mover offering the second mover the smallest feasible positive amount of money and the second mover accepting this offer. 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. This has been observed by Güth, Schmittberger, and Schwarze (1982), Hoffman and Spitzer (1985), Hoffman, McCabe, Shachat, and Smith (1994), and Bornstein and Yaniv (1998).[2] First movers may make generous proposals in ultimatum games because they have inequality-averse other-regarding preferences, as suggested by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), or altruistic other-regarding preferences, as suggested by Cox, Sadiraj, and Sadiraj (2002). 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. 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.[3]

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 (2003), which also studies the Punishment-MUG but observes different behavior. In order to identify the causes of the behavioral differences, we run an additional experimental treatment implementing the game with strategy method responses rather than sequential extensive form game responses. We also ask whether changing how the decision task is presented influences behavior by including a treatment in which the Punishment-MUG is framed as market exchange.

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 the second movers' 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's 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.[4]

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.

Given that our data differ from Trust game data reported by McCabe and Smith (2000), a systematic comparison of the studies provides more insight into positively reciprocal motivations. Specifically, we vary the cost of generous decisions by changing the payoff levels by 100%, thus allowing an examination of the price elasticity of generous behavior. We also vary the "social distance" inherent in the experimental protocol 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.

II. DISCRIMINATING AMONG MOTIVES IN THE PUNISHMENT-MUG GAME

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. The numbers at the end of a game tree branch are the dollar payoffs to the two subjects; the top number is the first mover's payoff and the bottom number is the second mover's payoff. For example, if the first mover chooses Take and the second mover chooses Tolerate, then the first mover receives $8 and the second mover receives $2.

The two punishment dictator control treatments shown in Figure 1 provide a way to discriminate among the alternative motivations. 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.

Mover 1 may choose Share rather than Take in the Punishment-MUG because of fear that, if he chooses Take, Mover 2 will choose Punish.[5] However, fear cannot explain the choice of "Share" in Punishment Control 1 because it is a dictator game. Similarly, Mover 2 may choose Punish rather than Tolerate to punish Mover 1 for attempting to obtain 80% of the total payoff for himself rather than making available the equal split.[6] However, punishment (of Mover 1) cannot explain Mover 2's choice of "Punish" in Punishment Control 2 because the decision node is determined randomly by the flip of a coin.[7]

III. PROCEDURES IN THE PUNISHMENT-MUG EXPERIMENT

We used a double-blind protocol similar to that used by Berg, Dickhaut and McCabe (1995) and Cox (2002; 2004). Double-blind payoff procedures ensure that the only person who knows the decision of a specific individual is the individual herself. A double blind payoff procedure provides a strong test for negative reciprocity because this procedure generates selfish behavior in standard dictator games but not in ultimatum games as in Camerer (2003).

All participants were undergraduate students who had not previously participated in a similar "fairness-game" experiment. 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.[8]

The sequence of events within a session was carefully controlled to ensure both subject understanding and privacy. Upon arrival for the experiment, each subject was paid the $5 participation fee and then entered the laboratory. The vacant stations between subjects and the raised partitions on three sides of the computer monitors provided significant privacy among subjects. After all participants had completed the computerized directions, a quiz was administered to test their comprehension.[9] 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 sealed key envelopes for implementing the double blind payoff procedure and handed out sheets of paper with written descriptions of the double-blind payoff procedures.[10] The description of the payoff procedures was read aloud to ensure that all participants were informed that everyone was receiving the same information.

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 second mover to determine nature's move. The procedure of flipping a coin has the advantages that subjects can verify the randomization process and the procedure is 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 decision 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.

As the amounts of money payoffs were determined by subjects' decisions, and appeared on the experimenters' monitor screen, money was inserted into envelopes that were in turn inserted into coded mailboxes in a room down a short hallway from the laboratory. After a total experiment time of approximately 45 minutes, subjects unlocked their mailboxes, retrieved their payoff envelopes, and exited.

IV. BEHAVIOR IN THE PUNISHMENT-MUG EXPERIMENT

Figure 1 presents data from the experiments with the Punishment-MUG and its motivational controls. Sixty-four percent of the first movers choose Take in the Punishment-MUG and 70% choose "Take" in Punishment Control 1, a difference that is not significant (p-value = 0.566 in a z-test / χ^2 test and 0.642 in a Fisher Exact test). Therefore, we conclude that first movers' behavior in the Punishment-MUG is not characterized by fear of negative reciprocity. This finding contrasts with Bolton and Zwick (1995, 115) who conclude that "second mover demands for fair treatment are responsible for ultimatum game outcomes." The percentage of first movers that we observe selecting Take is similar to that observed in the double blind "Zero Knowledge" treatment of their study (63%). However, the conclusions we draw differ because they compare behavior in a mini-ultimatum game not to choices in a dictator game but instead to decisions in an impunity game. In the impunity game a second mover's rejection of "Take" only lowers the second mover's payoff while rejection of "Share" yields $0 to both players.[11]

First movers' revealed expectations in the Punishment-MUG are consistent with second movers' behavior because only 21% of second movers choose Punish and thus Take results in an average payoff of $6.36 for first movers. While 21% of second movers in the Punishment-MUG did select Punish, this behavior cannot accurately be described as negative reciprocity. 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 two games is in the wrong direction to suggest negative reciprocity in the Punishment-MUG, and the difference is insignificant (p-value = 0.844 in a z-test / χ^2 test and 0.999 in a Fisher Exact test). Therefore we conclude that negative reciprocity does not explain behavior in this Punishment-MUG experiment. Second mover choices between ($0,$0) and ($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). In a similar game Güth, Huck, and Müller (2001) also found occasional rejections of equal split offers.

V. FURTHER EXPLORATION OF MOTIVES IN THE PUNISHMENT-MUG

The conclusion that negative reciprocity and fear of negative reciprocity do not characterize behavior is due in part to our observing behavior that is fundamentally different from results reported by Falk, Fehr, and Fishbacker (2003) for the Punishment-MUG. Based on the hypothesis tests presented in Table 1, our subjects were more likely to Take and less likely to Punish than their subjects.

It is important to consider how design differences between the two studies affect behavior. 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 30 new subject pairs 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 (second mover) 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.

Eighteen of the 30 first movers using the strategy method chose Take, which is not statistically different from the 64% who selected Take in our original sequential treatment (p-value = 0.718 in a z-test / χ^2 test and 0.819 in a Fisher Exact test). Also, 25 out of 30 second movers selected Tolerate, which is not statistically different from the 79% who chose Tolerate in our original sequential treatment (p-value = 0.566 in a z-test / χ^2 test and 0.764 in a Fisher Exact test). 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. 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. 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.[12]

Instead of playing a single game, in Falk, et al. subjects simultaneously made decisions for the Punishment-MUG and the three other games shown in Figure 2, 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). 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 2. This is the pattern of responses in their within-subjects treatment.

Other differences between the studies include the use of single blind vs. double blind payoff procedures, cultural differences in the subject pools, and the real payoff levels as subjects were paid in different currencies. However, 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. This suggests 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 embedded. 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.

Hoffman, et al. (1994) found that a buyer-seller framework led to more materially self-regarding behavior in ultimatum games. Given our focus on the conditions under which we observe reciprocity and that we have established that mini-ultimatum games can differ from standard ultimatum games, we conducted a market-framed Punishment-MUG with 30 new subject pairs. In this version of the game Mover 1 is a seller who sets a price for a good with zero marginal cost. Mover 2 is a buyer who values the good at \$10 and 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. The same experimental procedures were employed as before except for the necessary changes in directions and computer interface.[13] Seventeen of the thirty sellers set a price of \$8. This behavior is similar to that observed in our original sequential Punishment-MUG treatment (p-value = 0.503 in a z-test / χ^2 test and 0.647 in a Fisher Exact test). A surprising feature of the data is that a higher percentage of unequal offers were vetoed (or punished) in the market frame (seven of 17) than in the sequential treatment (eight of 39). This is in the opposite direction than predicted based upon the results of Hoffman, et al. (1994); however, this difference is not significant (p-value = 0.108 in a z-test / χ^2 test and 0.188 in a Fisher Exact test). Only one of the 13 sellers who set a price of \$5 did not complete a sale. Thus we conclude that second mover preferences in the Punishment-MUG are fairly robust when the player considers this game in isolation. Ultimately, the only evidence of negative reciprocity that we find occurs in the Falk, et al. experiment when the decision is embedded in a larger context.

VI. DISCRIMINATING AMOING MOTIVES IN THE TRUST GAME

In the Trust game presented in Figure 3, 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. The two trust dictator control treatments shown in Figure 3 provide a way to discriminate among alternative motivations.

Mover 1 may choose Engage rather than Exit in the Trust game because she trusts that Mover 2 will not Defect.[14] Trust Control 1 is a dictator game in which Mover 1 chooses between (\$5,\$5) and (\$0,\$20). 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 1 choose (\$5,\$5), the difference between first movers' decisions in the Trust game and Trust Control 1 discriminates between choices of Engage that are motivated by trust in positive reciprocity and those that are motivated by non-reciprocal other-regarding preferences. If choices of (\$0,\$20) were to be observed in Trust Control 1, then it would be necessary to experiment with an additional trust control in order to discriminate between first-mover motives in the Trust game.

Mover 2 may choose Cooperate rather than Defect in the Trust game because of a social norm for positive reciprocity.[15] 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. For a choice of Cooperate in the Trust game to be interpreted as reciprocity, it must be known that Mover 2 prefers the outcome ($0,$20) to the outcome ($7.50,$12.50) in the absence of a prior action by Mover 1. Comparing behavior differences between Trust Control 2 and Trust distinguishes behavior motivated by reciprocity from behavior motivated by non-reciprocal other-regarding preferences.

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

VII. BEHAVIOR IN THE TRUST GAME EXPERIMENTS

The results from the Trust game and its controls are reported in Figure 3. Only four of the 30 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 higher proportion of "second movers" chose "Cooperate" in the dictator control treatment than in the Trust game but the difference is not statistically significant (p-value = 0.480 in a z-test / χ^2 test and 0.529 in a Fisher Exact test). This result is clearly inconsistent with reciprocity being the motivation for subjects' choice of Cooperate in the Trust game. We conclude that positive reciprocity does not explain behavior in this extensive form Trust game experiment.

It is worth noting that the behavior we observe in Trust Control 2 is consistent with what we observed in Punishment Control 1. Our third main finding is that only two of 30 Trust Control 1 dictators selected the ($0, $20) outcome over ($5, $5), whereas 17 of 30 of the first movers in the Trust game chose Engage. Observed differences between Trust Control 1 and the Trust game provide evidence that, in the Trust game, first mover choices of Trust are, in fact, motivated by trust (p-value < 0.001 in both a z-test / χ^2 test and a Fisher Exact test). Consistent with a growing literature[16] on first movers' failure to correctly anticipate second movers' defection trust is misplaced because Engage yielded an average payoff of $1.80 to first movers whereas the choice of Exit yielded a certain payoff of $5.

VIII. 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), as shown in Table 2.[17] Again, it is important to discern how design differences affect reciprocal behavior.

These experiments differed in two dimensions. First, the payoffs in their experiment were double those of our Trust game shown in Figure 3. Second, their experimental protocol had less "social-distance" between the subjects and the experimenters and among the subjects. Specifically, they used a single-blind payoff procedure with exactly 6 subject pairs per session whereas we used a double blind payoff procedure and had between 7 and 10 subject pairs per session. We do not consider number of subjects and payoff procedures separately because both features could make their subjects more concerned about how others might judge their behavior.

To identify which design feature caused the behavioral change, we conducted an additional treatment with 27 new subject pairs after first replicating McCabe and Smith's results.[18] In this new Trust game treatment, the payoffs were double those shown in Figure 3 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, ten of the 14 second movers who had a decision to make chose Defect. Table 3 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, it remains to be determined if reciprocity exists when social distance is low. 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. An additional treatment was conducted with 24 new subject pairs playing Trust Control 2 for twice the payoffs shown in Figure 3, 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 3 for "second movers" in the original Trust Control 2, 33% in each treatment. Based on this we conclude that reciprocity does exist under low social distance protocols (p-value = 0.015 in a z-test / χ^2 test and 0.023 in a Fisher Exact test). Recent work by McCabe, Rigdon, and Smith (2003) draws the same conclusion using data from a modified version of the trust game.

Since behavior in Trust Control 2 was the same with our two payoff protocols, we conclude that behavior motivated by non-reciprocal other-regarding preferences is unchanged by altering the level of social distance. This finding also suggests that the double blind payoff protocol 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. 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 in Trust Control 2, we conclude that trust is present in the low social distance environment as well, and is a robust behavioral phenomenon.

From these six 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.[19]

IX. 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. 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.

Negative reciprocity was examined via sequential play of the Punishment-MUG using a double-blind protocol in which subjects' decisions were anonymous both to other subjects and to the experimenters. Behavior did not differ between the Punishment-MUG and the dictator control in which there was no possibility of negative reciprocity; approximately two-thirds of the first movers opted for the self-favoring unequal split. Rejection of the unequal proposal in the Punishment-MUG was similar to the rejection rate for the control game where a coin flip generated the unequal split, 21% and 23% respectively. Thus, fear of negative reciprocity was not observed in the behavior of first movers and negative reciprocity was not observed in the behavior of second movers.

The above-described behavior is significantly different than Punishment-MUG behavior in experiments reported by Falk, Fehr, and Fischbacher (2003), who did observe significant negative reciprocity. Falk, et al. used strategy responses in which the "second mover" entered decisions for all possible nodes in a game without observing the first mover's decision. However, when we conducted experiments with the Punishment-Mug using strategy responses the behavior was statistically indistinguishable from our original treatment and thus we conclude that the differences in observed levels of negatively-reciprocal behavior cannot be attributed to effects of extensive-form game-playing. Based on our findings, it appears that the feature of the Falk, et al. design that may account for significant negatively-reciprocal behavior is the simultaneous play of four games. 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) also provide support for this interpretation. Thus we conclude that negative reciprocity is contingent on the larger context in which the decision task is embedded. In a similar vein, we alter the context of the game by framing the decision as market exchange to explore the robustness of our finding. The data from the market-exchange treatment were not significantly different from our original treatment; hence the absence of negatively-reciprocal behavior in the Punishment-MUG is robust.

Positive reciprocity was examined in the Trust game and appropriate dictator controls using a double-blind protocol. We found that first movers' behavior was characterized by significant trust in positive reciprocity. However, there was no evidence of positive reciprocity as there was not a significant difference between second movers' play in the Trust game and play by "second movers" in the control treatment.

The behavior we observed is in direct contrast to work previously reported by McCabe and Smith (2000). The design differences between the two studies include the social distance of the protocol and our use of payoffs that were one-half the payoffs used by McCabe and Smith. After replicating their results, we ran another treatment using the higher payoff level and our protocol for the Trust game. Data from this treatment were not significantly different from data in our initial Trust game results but were significantly different from the findings of McCabe and Smith and our replication. 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 some findings by Hoffman, et al. (1994).[20] Charness, et al. (2001) also report that social distance affects behavior.[21]

To determine if behavior in the low social distance protocol is motivated by reciprocity, we conducted an additional second-mover control treatment with the high payoff and low –social distance protocol. Data from this implementation were not significantly different from our initial second-mover control game, but the data were significantly different from the low-social-distance, high-payoff Trust game; that is, we observed significant positive reciprocity in this situation. Thus we conclude 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.

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. Our triadic design discriminates between behavior motivated by trust or reciprocity and behavior motivated by (intentions-unconditional) other-regarding preferences characterized by altruism as described by Cox, Sadiraj, and Sadiraj (2002), maxi-min considerations of Charness and Rabin (2002), or inequality aversion as developed by Fehr and Schmidt (1999) and 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.

References

Berg, J., J. Dickhaut, and K. McCabe. "Trust, Reciprocity and Social History." *Games and Economic Behavior,* 10 (1), 1995, 122-42.

Bolton, G. E., J. Brandts, and A. Ockenfels. "Fair Procedures Evidence from Games Involving Lotteries." Working Paper, Penn State University, July 2000.

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

Bolton, G. E., and R. Zwick. "Anonymity versus Punishment in Ultimatum Bargaining." *Games and Economic Behavior*, 10 (1), 1995, 95-121.

Bornstein, G., and I. Yaniv. "Individual and Group Behavior in the "Ultimatum Game: Are Groups More 'Rational' Players?" *Experimental Economics*, 1(1), 1998, 101-08.

Camerer, C. *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton: Princeton University Press, 2003.

Cason, T., and D. Friedman. "Buyer Search and Price Dispersion: A Laboratory Study." Discussion Paper, University of California at Santa Cruz, July 2001.

Charness, G., E. Haruvy, and D. Sonsino. "Social Distance and Reciprocity: The Internet vs. the Laboratory." Discussion Paper, University of California at Santa Barbara, 2001.

Charness, G., and M. Rabin. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics.* 117(3), 2002, 817-69.

Cherry, T., P. Frykblom, and J. Shogren. "Hardnose the Dictator." *American Economic Review*, 92(4), 2002, 1218-21.

Cooper, D. J., and J. B. Van Huyck. "Evidence on the Equivalence of the Strategic and Extensive Form Representation of Games." *Journal of Economic Theory*, 110 (2), 2003, 290-308.

Cox, J. C. "Trust, Reciprocity, and Other-Regarding Preferences: Groups vs. Individuals and Males vs. Females," in *Advances in Experimental Business Research*, edited by R. Zwick and A. Rapoport. Boston: Kluwer Academic Publishers, 2002.

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

Cox, J. C., R. M. Isaac, P. Cech, and D. Conn. "Moral Hazard and Adverse Selection in Procurement Contracting." *Games and Economic Behavior*, 17 (2), 1996, 147-76.

Cox, J. C., and R. L. Oaxaca. "Laboratory Experiments with a Finite Horizon Job Search Model." *Journal of Risk and Uncertainty*, 2(3), 1989, 301-29.

______, and ______. "Is Bidding Behavior Consistent with Bidding Theory for Private Value Auctions?" in *Research in Experimental Economics* VI, edited by R. M. Isaac. Greenwich: JAI Press, 1996.

______, and ______. "Good News and Bad News: Search from Unknown Wage Offer Distributions." *Experimental Economics*, 2(3), 2000, 197-225.

Cox, J. C.,, K. Sadiraj, and V. Sadiraj. "Implications of Trust, Fear, and Reciprocity for Modeling Economic Behavior." Working Paper, University of Arizona, May 2002, revised 2004.

______, ______, and ______. "A Theory of Competition and Fairness for Egocentric Altruists." Working Paper, University of Arizona, June 2002.

Davis, D., and C. Holt. *Experimental Economics.* Princeton: Princeton University Press, 1993.

Deck, C. A. "A Test of Behavioral and Game Theoretic Models of Play in Exchange and Insurance Environments." *American Economic Review*, 91(5), 2001, 1546-55.

Falk, A., E. Fehr, and U. Fischbacher. “On the Nature of Fair Behavior.” *Economic Inquiry*, 41(1), 2003, 20-6.

Fehr, E., and S. Gächter. “Fairness and Retaliation: The Economics of Reciprocity.” *Journal of Economic Perspectives.* 14(3), 2000, 159- 81.

Fehr, E., and K. M. Schmidt. “A Theory of Fairness, Competition and Cooperation.” *Quarterly Journal of Economics*, 114(3), 1999, 817-68.

Forsythe, R., J. Horowitz, N. Salvin, and M. Sefton. “Fairness in Simple Barging Experiments.” *Games and Economic Behavior*, 6(3), 1994, 347-69.

Güth, W., S. Huck, and W. Müller. “The Relevance of Equal Splits in Ultimatum Games.” *Games and Economic Behavior*, 37(1), 2001, 161-9.

Güth, W., R. Schmittberger, and B. Schwarze. “An Experimental Analysis of Ultimatum Bargaining.” *Journal of Economic Behavior and Organization*, 3(4), 1982, 367-88.

Harrison, G. W., and P. Morgan. “Search Intensity in Experiments.” *Economic Journal*, 100(401), 1990, 478-86.

Hoffman, E., K. A. McCabe, K. Shachat, and V. L. Smith. “Preferences, Property Rights, and Anonymity in Bargaining Games.” *Games and Economic Behavior*, 7(3), 1994, 346-80.

Hoffman, E., and M. L. Spitzer. “Entitlement, Rights and Fairness: An Experimental Examination of Subjects’ Concepts of Distributive Justice.” *Journal of Legal Studies*, 14(2), 1985, 259-97.

McCabe, K. A., and V. L. Smith. “A Comparison of Naïve and Sophisticated Subject Behavior with Game Theoretic Predictions.” *Proceedings of the National Academy of Sciences*, 97 (7), 2000, 3777-81.

McCabe, K. A., M. L. Rigdon, and V. L. Smith. “Positive Reciprocity and Intentions in the Trust Game.” *Journal of Economic Behavior and Organization*, 52(2), 2003, 267-75.

McKelvey, R. D., and T. R. Palfrey. “An Experimental Study of the Centipede Game.” *Econometrica*, 60(4), 1992, 803-36.

______, and ______. “Quantal Response Equilibria for Extensive Form Games.” *Experimental Economics*, 1(1), 1998, 9-41.

Rabin, M. “Incorporating Fairness into Game Theory and Economics.” *American Economic Review*, 83(5), 1993, 1281-302.

Ruffle, B. “More is Better, But Fair is Fair: Tipping in Dictator and Ultimatum Games.” *Games and Economic Behavior*, 23(2), 1998, 247-65.

Smith, V. L. “Microeconomic Systems as an Experimental Science” *American Economic Review*, 72(5), 1982, 923-55.

Sonnegard, J. “Determination of First Movers in Sequential Bargaining Games: An Experimental Study.” *Journal of Economic Psychology*, 17(3), 1996, 359-86.

Footnotes

* 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.

Cox: Arizona Public Service Professor, Department of Economics and Director, Economic Science Laboratory, University of Arizona, Tucson, AZ 85721-0108. Phone 1-520-621-2164, Fax 1-520-621-8450, E-mail jcox@eller.arizozna.edu

Deck: Assistant Professor, Department of Economics, University of Arkansas, Fayetteville, AR 72701-1201. Phone 1-479-575-6226, Fax 1-478-575-3241, E-mail cdeck@walton.uark.edu

[1] Some of this literature is reviewed in Fehr and Gächter (2000). It is not the case that behavior in games is never consistent with self-regarding preferences, but rather that behavior depends on perceived property rights as shown by Sonnegard (1996) and Ruffle (1998). Cherry, Frykblom, and Shogren (2002) find significant levels of self-regarding behavior in a dictator game where people had to earn their money.

[2] Camerer (2003) reports that the ultimatum game is also robust to the double blind payoff procedure.

[3] 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. See Cox (2002; 2004) for a discussion of the triadic design and a model that represents the different effects on subjects' motivations with state-dependent utility functions.

[4] Some other explanations for this behavior have been developed, including McKelvy and Palfrey's (1992; 1998) development of the game-theoretic implications of stochastic decision-

making in the form of quantal response and Rabin's (1993) model that incorporates intentions. 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.

[5] Alternatively, Mover 1 may choose Share due altruistic or inequality-averse other-regarding preferences.

[6] Alternatively, Mover 2 may Punish due inequality-averse preferences.

[7] The random procedure itself might influence second mover choices. Bolton et al. (2000) investigate the impact of random procedures on experimental subjects. They find that individuals are more likely to reject unfavorable lopsided offers in a symmetric ultimatum game if the randomization process is biased in the other player's favor. An alternative control for second movers' motives would be a simple dictator game in which a second mover chooses between ($8,$2) and ($0,$0). However, as Falk, Fehr, and Fischbacher (2003) and Güth, Huck, and Müller (2001) have shown, behavior in mini-ultimatum games changes dramatically depending on the payoffs associated with unreached decision nodes. Therefore, using this alternative control would confound the effects of the first mover's action with the elimination of a potential payoff pair.

[8] There was no particular order to these treatments. It was not the case that all of the Punishment-MUG sessions were completed before the control treatments were begun or vice versa. Similarly, there was no particular order to the Trust game and its dictator controls either. Of course the additional treatments that were designed to identify how design changes affect behavior were conducted after the initial treatments were completed.

[9] 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 again with the quiz and no significant difference between the two data sets was found. Therefore, re-running the control experiments was considered unnecessary.

[10] The subject instructions, quiz, and description of the double-blind payoff procedures are available from the authors upon request.

[11] The impunity game fundamentally alters the decision context by making opting for an equal split more risky and perhaps changes the subject's frame of reference. Various other protocol differences may also help explain the difference in behavior, including their use of a single blind procedure in the impunity game, the fact that subjects played a series of games, and the relatively low stakes of their games.

[12] An alternative approach to the strategy method would be to have each individual make choices as both first mover and second mover and then let Nature determine the actual role of a subject. While this method might lead subjects to reflect more on their motives, it would differ from sequential play where subjects are in only one role.

[13] 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.

[14] Alternatively, a first mover may choose Engage because of altruistic preferences, preferring both ($7.50,$12.50) and ($0,$20) to ($5,$5), due to a preference for efficiency, as in Charness and Rabin (2002), or due to a preference to let the other person make a choice, as in Cooper and Van Huyck (2003).

[15] Alternatively, a second mover may choose Cooperate due to inequality-averse or altruistic other-regarding preferences.

[16] See Camerer (2003)

[17] Our results also differ from McCabe, Rigdon and Smith (2003), who compare behavior between a Trust game and a second mover control with a single blind protocol. They conclude that reciprocity does explain behavior. While the behavior they report is similar to McCabe and Smith (2000), the payoff structure is different. Specifically, defection lowers efficiency and does not lead to the same extreme payoff inequality.

[18] The high cooperation rate reported by McCabe and Smith (2000) was based on the responses of only twelve individuals. Therefore, we first attempted 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. Also, their single-blind procedure of calling subjects by name to come and individually receive payoffs directly from the experimenters was used. Table 2 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.

[19] 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 in Cox (2002; 2004) and the moonlighting game in Cox, Sadiraj, and Saditraj (2004). One notable difference between those games and the Trust game is that the former have many more possible

responses that the subjects can make. A denser message space may allow decision-makers to undertake less costly reciprocal behavior.

[20] Hoffman, McCabe, Shachat, and Smith (1994) find that single-blind and double-blind payoff protocols can elicit significantly different behavior in some contexts. Interestingly, Hoffman, et al. found that social distance did influence behavior in a dictator game. However, they did not conduct both single blind and double blind dictator games using their "random entitlement" procedure. Instead they relied on data from Forsythe et al. (1994). The apparent discrepancy could be due to procedural differences between those two studies or to the use of more vs. less dense message spaces.

[21] Charness, Haruvy, and Sonsino (2001) 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. 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. They found that the level of social distance had a significant effect on behavior.

Captions and Notes for Figures

FIGURE 1

Behavior in Punishment-MUG and Dictator Controls

Notes: Numbers beside each move indicate the number of subjects selecting that move. N denotes a move by nature. The numbers beside nature's decision branches indicate the probability of the move being selected by a coin flip.

FIGURE 2

The Experiments of Falk, Fehr, and Fischbacher (2003)

FIGURE 3

Behavior in Trust and Dictator Controls

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

TABLE 1

Comparison of Punishment-MUG Results with Falk, et al. (2003) Data

		Take	Punish	Accept
Falk, et al.	Number of Subjects Selecting Action	14	20	45
	Total Subjects Facing Decision	45	45	45
Punishment-MUG	Number of Subjects Selecting Action	39	8	21
	Total Subjects Facing Decision	61	39	22
p-value for z-test / χ^2 test		<0.001	0.020	0.150
p-value for Fisher Exact test		0.002	0.036	0.328

Notes: The null hypothesis is Ho: there is no difference in behavior between the two data sets. p-values are for the two sided alternative.

TABLE 2

Comparison of Trust Game Results with McCabe and Smith (2000) Data

	Trust game		McCabe and Smith Data		Replication of McCabe and Smith	
	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

Data Sets Being Compared	Decision	p-value for z-test / χ^2 test	p-value for Fisher Exact test
Trust vs. McCabe and Smith Data	Engage	0.624	0.784
	Cooperate	0.006	0.010
Trust vs. Replication of McCabe and Smith	Engage	0.862	>0.9
	Cooperate	0.035	0.061
Replication of McCabe and Smith vs. McCabe and Smith Data	Engage	0.777	>0.9
	Cooperate	0.471	0.673

Notes: The null hypothesis is Ho: there is no difference in behavior between the two data sets. p-values are for the two sided alternative.

TABLE 3

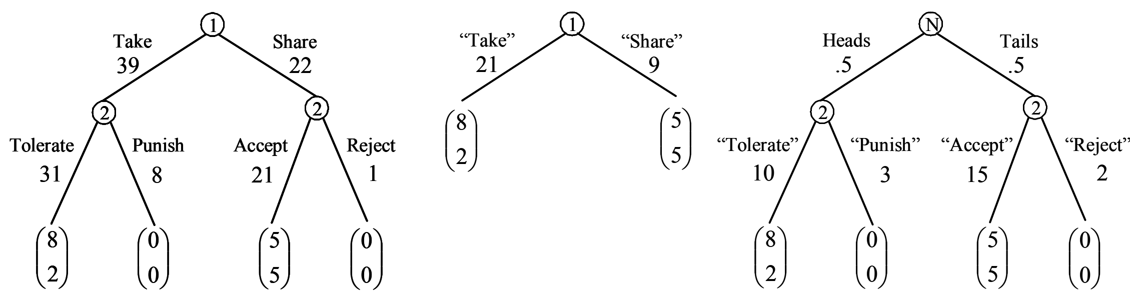
Effects of Social Distance and Payoff Level on Behavior in the Trust Game

	High Social Distance Half Payoff (Trust game)		Low Social Distance Full Payoff (McCabe and Smith 2000 and replication)		High Social Distance Full Payoff	
	Observed Frequency	Sample Size	Observed Frequency	Sample Size	Observed Frequency	Sample Size
Engage	17	30	25	48	14	27
Cooperate	4	17	17	25	4	14

Environments Being Compared	Decision	p-value for z-test / χ^2 test	p-value for Fisher Exact test
High Social Distance Half Payoff vs. Low Social Distance Full Payoff	Engage	0.689	0.816
	Cooperate	0.005	0.011
High Social Distance Half Payoff vs. High Social Distance Full Payoff	Engage	0.718	0.793
	Cooperate	0.749	>0.9
Low Social Distance Full Payoff vs. High Social Distance Full Payoff	Engage	0.985	>0.9
	Cooperate	0.012	0.024

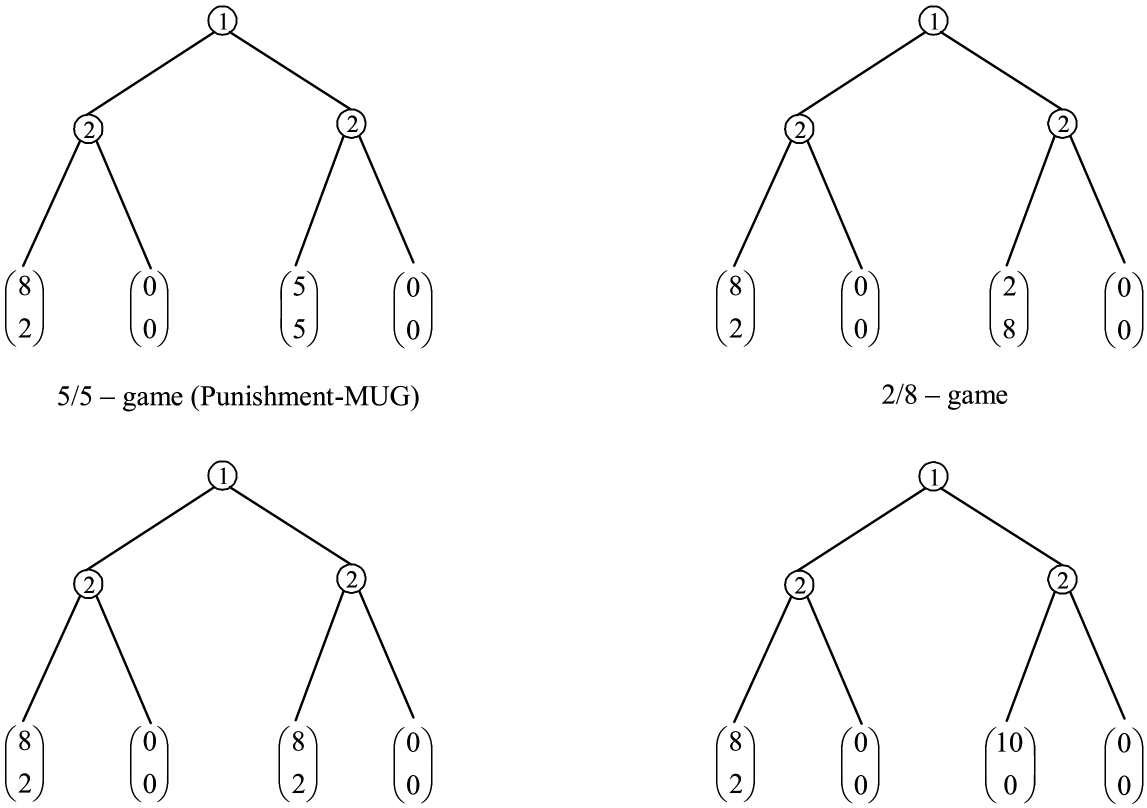
Notes: The null hypothesis is Ho: there is no difference in behavior between the two data sets. p-values are for the two sided alternative.

FIGURE 1



Punishment-MUG Punishment Control 1 Punishment Control 2

FIGURE 2



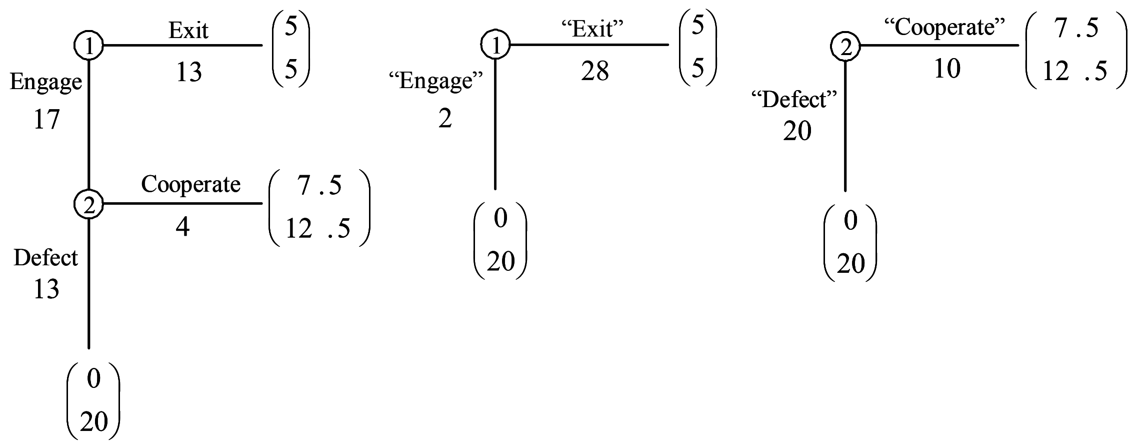
5/5 – game (Punishment-MUG)

2/8 – game

8/2 – game

10/0 – game

FIGURE 3



Trust　　Trust Control 1　　Trust Control 2