Essays on Political and Experimental Economics

Sadiraj, V.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 6

Trust, Fear, Reciprocity, and Altruism

6.1 Introduction

Economic and game-theoretic modeling has historically focused on the model of “self-regarding preferences” in which agents are assumed to be exclusively concerned with maximizing their own material payoffs. This model predicts behavior quite well in many types of controlled experiments. But there is now a large body of experimental literature that has produced replicable patterns of inconsistencies with the self-regarding preference model’s predictions in contexts involving salient fairness considerations or opportunities for cooperation. This literature was reviewed in a recent survey paper on “the economics of reciprocity” by Fehr and Gächter (2000a).

In order to obtain data that can guide development of economic and game-theoretic models, we need to be able to discriminate between actions with alternative motivations. In this chapter, we use a triadic experimental design to discriminate between actions motivated by preferences over outcomes and actions motivated by attributions of intentions in the moonlighting game (see Abbink, Irlenbusch, and Renner (2000)). In the moonlighting game, a first-mover can either give money to a paired second mover or take money from the second mover. Any amount given is tripled by the experimenter. Any amount taken is

---

This chapter draws heavily from Cox, Sadiraj and Sadiraj (2001a).
not transformed by the experimenter. After a second mover learns of the tripled amount sent or the amount taken by the paired first mover, the second mover has an opportunity to give or take money from the first mover. Each dollar that the second mover gives to the first mover costs the second mover one dollar. Each dollar that the second mover takes from the first mover costs the second mover 33 cents. Our concern is what motivations for actions can be inferred from observations in the moonlighting game. A first mover may give money to a second mover because of altruistic other-regarding preferences. Alternatively, a first mover may give money to the paired second mover because of trust that the second mover will return part of the profit from the experimenter’s tripling of the amount sent. Furthermore, a first mover may refrain from taking money from the paired second mover because of altruistic or inequality-averse other-regarding preferences. Alternatively, a first mover may refrain from taking money from the paired second mover because of fear that the second mover will retaliate by subsequently taking money from him. Turning now to the second mover, we note that a second mover may return part of the tripled amount sent by the paired first mover because of positive reciprocity, by which we mean a motivation to adopt a costly action to benefit someone whose intentional behavior has benefited oneself. Alternatively, the second mover may return part of the tripled amount sent because of altruistic or inequality-averse other-regarding preferences. Next consider possible second-mover motivations for incurring the cost of taking money from a first mover that took money from the second mover. The second mover may take money from the first mover because of negative reciprocity, by which we mean a motivation to adopt a costly action to inflict harm on someone whose intentional behavior has inflicted harm on oneself. Alternatively, the second mover may take money from the first mover because of inequality-averse other-regarding preferences.

The observations in the moonlighting game reported by Abbink, Irlenbusch, and Renner (2000) do not discriminate between actions with alternative motivations. This is our reason for incorporating the moonlighting game into a triadic experimental design that includes dictator control treatments that make these discriminations possible.

This chapter is organized as follows. In Section 6.2, we give our definitions of trust, positive direct reciprocity, fear of negative reciprocity and negative reciprocity. In Section
6.2. DEFINITIONS

6.3, we present the experimental design which includes the moonlighting game and two dictator games. The procedures of the experiments are reported in this section as well. In Section 6.4, we report and analyze the behaviour of the subjects in the experiment. In Section 6.5, we state our conclusions followed by some remarks. Appendix C contains a translation of the instructions given to the experiment subjects in the moonlighting game and two dictator games.

6.2 Definitions

The concepts of trust and positive reciprocity used in this chapter are defined as follows.

Definition 6.1 Agent 1 undertakes an action that exhibits trust if the chosen action:

(a) creates a monetary gain that could be shared with agent 2; and

(b) exposes agent 1 to the risk of a loss of utility if agent 2 defects and appropriates too much of the monetary gain.

Definition 6.2 Agent 2 undertakes an action that exhibits positive direct reciprocity if the chosen action:

(a) follows a trusting action by agent 1;

(b) gives agent 1 a monetary gain; and

(c) is undertaken instead of an available alternative action that would produce outcomes preferred by agent 2 in the absence of the trusting action by agent 1.

The concepts of fear and negative reciprocity used in this paper are defined as follows.

Definition 6.3 Agent 1 undertakes an action that exhibits fear of negative reciprocity if, in two otherwise-identical environments, he:

(a) foregoes an opportunity to take money from agent 2 when agent 2 has an opportunity to retaliate; and
(b) takes money from agent 2 when agent 2 does not have an opportunity to retaliate.

Definition 6.4 Agent 2 undertakes an action that exhibits **negative direct reciprocity** if in two otherwise-identical environments in which agent 1 has been allocated the higher monetary payoff he:

(a) reduces both agents’ monetary payoffs when agent 1 is responsible for the unfavorable allocation; and

(b) does not reduce their monetary payoffs when agent 1 is not responsible for the unfavorable allocation.

Note that the above definitions of *observable* positive and negative reciprocity incorporate a possible dependence of the *inferred* preferences over outcomes on the process that generates those outcomes and attributions of the intentions of others. And the definitions of *observable* trust and fear incorporate a possible dependence of the *inferred* motivations on the process that generates the outcomes. The triadic experimental design explained in Section 6.3 makes it possible to discriminate between the implications of other-regarding preferences and trust, fear, or reciprocity.

### 6.3 Experimental design and procedures

The experimental design includes the moonlighting game and two specially-designed dictator games.

#### 6.3.1 The three games

The moonlighting game is experimental treatment $A$. In treatment $A$, the initial endowment of each individual is 10 euros. A first mover chooses an amount, $s^a$ to send to the second mover, where $s^a \in S$ and

$$ S = \{-5, -4, \ldots, 0, 1, \ldots, 10\}. $$
If $s^a < 0$ then the first mover takes money from the second mover. If $s^a > 0$ then the first mover gives money to the second mover. Thus, if $s^a < 0$ then the second mover's money payoff is decreased by the amount $-s^a$ and the first mover's money payoff is increased by the amount $-s^a$. If $s^a > 0$ then the second mover's money payoff is increased by $3s^a$ and the first mover's money payoff is decreased by $s^a$. If $s^a = 0$ then no money payoff is changed.

The choice of $s^a$ by the first mover selects the $\Gamma(s^a)$ subgame, in which the second mover chooses $r^a \in R(s^a)$, where

\[
R(s^a) = \{m(s^a), m(s^a) + 1, \ldots, M(s^a)\}.
\]

If $r^a < 0$ then the first mover's money payoff is decreased by $-3r^a$ and the second mover's money payoff is decreased by $-r^a$. If $r^a > 0$ then the first mover's money payoff is increased by $r^a$ and the second mover's money payoff is decreased by $r^a$.

The second mover is not allowed to choose a return amount that would leave the first mover with a negative monetary payoff; therefore, $r^a \geq -((10 - s^a)/3)$. Also, $r^a$ must be an integer; therefore,

\[
m(s^a) = -\max \{y \in \mathbb{Z} \mid y \leq ((10 - s^a)/3)\}
\]

where $\mathbb{Z}$ is the set of integers. The second mover is not allowed to choose a return amount that would leave herself with a negative payoff; therefore $r^a \leq M(s^a)$, where

\[
M(s^a) = \begin{cases} 
10 + s^a, & \text{if } s^a < 0 \\
10 + 3s^a, & \text{otherwise.}
\end{cases}
\]

The set $S$ of feasible choices for a first mover was chosen because it is a natural extension of the set of feasible choices in the investment game. The set of integers weakly between 0 and 10 is the feasible set of first-mover choices in the Berg, Dickhaut, and McCabe (1995) investment game experiment. Our choice of $S$ for the moonlighting game\footnote{In Abbink, Irlenbusch, and Renner (2000) first movers could send an amount that varied from -6 to +6.}

11
extends the set of feasible choices to include the possibility of taking any integer amount up to 5, which is one-half of the second mover’s endowment. Thus we can address the question of the effect on first-mover decisions of adding the opportunity to take money by comparing our data to the Berg et al. data.

Limiting the amount that can be taken by a first mover to be no more than 5 preserves the ability of a second mover to retaliate if he chooses to. For example, if a first mover takes the maximum of 5, then the second mover has the option of paying a cost in integer amounts up to all of his remaining balance of 5. If a first mover takes 5 from the second mover, then he has 15 before the second mover’s decision. The second mover then has the option of paying all of his remaining 5 to reduce the first mover’s credit balance to zero. The feasible choice sets for the other treatments are as follows.

Treatment B is a dictator game with the same strategy set for the “first mover” as in treatment A, the moonlighting game. Thus, the “first mover” chooses \( s^b \in S \), where \( S \) is defined in statement 6.1. The “second mover” has no decision to make.

Treatment C is a dictator game in which the “first mover” has no decision and the “second mover” has the same strategy set that he has in the subgame selected in treatment A. Thus, the “second mover” chooses \( r^c \in R(s^a) \), where \( R(s^a) \) is defined in statement 6.2.

6.3.2 Procedures

The experiment sessions were run with custom computer software in the CREED laboratory at the University of Amsterdam in the fall of 2000. The subjects assembled in a sign-in room. They registered on a subject list and picked up copies of printed instructions from a stack on a table. The subjects drew small envelopes and folded “notes” from two different boxes, each containing items that were identical on the outside. Each envelope contained a mailbox key with a unique identification code. The subjects were asked not to open their envelopes until they were seated at computers in the laboratory. The key codes were to be used for subject identification for money payoffs. One-half of the notes contained the symbol # and one-half contained the symbol *. The random assignment of symbols on the notes implemented the random assignment of subjects to the two sections.
6.3. EXPERIMENTAL DESIGN AND PROCEDURES

of the laboratory.

Subsequently, the subjects walked a few feet down the hallway and entered the laboratory through either the door marked with # or the door marked with *. The experimenter stood in the hallway well back from the two doors, and in a position where observation of which subject approached which computer was impossible. After all subjects had entered the laboratory, the doors were closed for the duration of the experiment. The laboratory was divided into two sections by a floor to ceiling partition. One section was accessed through the door marked # and the other section was accessed through the door marked *. The windows between the experimenters' control room and one of the two parts of the laboratory were completely covered by blinds. Thus, the two groups of subjects had no verbal, visual, or other contact with each other or with the experimenters during the decision-making part of the experiment.

The subjects read the instructions on their computer monitors. Printed copies of the instructions were made available in case the subjects wanted to review the instructions during the decision-making part of the experiment. The instructions referred to the subjects only as being in group X or group Y. Terms such as first mover, second mover, proposer, responder, etc. were avoided. The instructions stated that subjects could "increase" or "decrease" their own and the paired subject's "account balances". The instructions did not use the words "send" and "return" for the amounts transferred by first and second movers. Other, possibly more evocative verbs, such as "give", "take", "reward", and "punish" were avoided. Tables in the instructions presented all feasible actions and their consequences for both subjects in a pair of first and second movers.

The end of the instructions directed the subjects to enter their key codes into their computers and then proceed to answer the questions that would appear on their computer monitors. The questions were intended to test subjects' understanding of the experimental tasks and procedures. If a subject answered a question incorrectly, she was informed of this and then asked to try again. After all of the subjects answered all questions correctly, the decision-making part of the experiment began.

The decision-making part of the experiment proceeded as follows. First, the monitor computer randomly determined which room, # or * was the room with group X subjects
and which was the room with group Y subjects. The pairing of group X and group Y subjects was established by where the subjects sat in the two separated parts of the laboratory. Thus the subjects had no way of knowing who they were paired with. And the experimenters had no way of knowing which subject sat at which computer. Salient payoffs were possible because the subjects entered their key codes in their computers. The payoff procedure was double blind: (i) subject responses were identified only by the key codes that were private information of the subjects; and (ii) money payoffs were collected in private from sealed envelopes contained in coded mailboxes.

In treatment A, each individual in each group was credited with 10 euros. At the time of the experiment, one euro was worth a little less than one dollar. Each individual in group X was given the opportunity to give or take money from an anonymously-paired person in group Y. Thus, a group X subject could choose to neither give nor take. He could choose to give any integer amount up to all 10 of his euros to the group Y subject. Or he could take any integer amount up to 5 euros from the group Y subject. Any amount given was tripled by the experimenters. Any amount taken was not transformed by the experimenters. Subsequently, each individual in group Y was given the opportunity to give or take money in integer amounts from the anonymously-paired person in group X. Each euros given to the group X subject cost the group Y subject one euro. The amount given could not exceed the amount that would leave the group Y subject with zero euros. Each three euros taken from the group X subject cost the group Y subject one euro. The amount taken could not exceed the amount that would leave the group X subject with zero euros.

Treatment B differed from treatment A only in that the individuals in group Y had no decision to make. Thus, treatment B was a dictator experiment in which group X subjects had the same strategy set as in treatment A.

Treatment C differed from treatment A as follows. Individuals in group X had no decision to make. Each subject pair, \( j = 1, 2, \ldots, 30 \), was informed that the person in group X had a beginning account balance of \( 10 + A_j \) and the person in group Y had a beginning account balance of \( 10 + B_j \). A subject pair in treatment C was informed of the amounts, but not told that they had been determined by the decision of the group
6.4. **SUBJECTS' BEHAVIOR IN THE EXPERIMENT**

The decision to withhold this information was based on the judgment that it might motivate indirect reciprocity by the subjects, which would be inappropriate in this control treatment. A desire to avoid an alternative type of indirect reciprocity also accounts for the way the endowments were implemented. A different procedure than we used would be to first endow each subject in every pair with 10 euros and, subsequently, have the experimenter or another third party alter the endowments for pair \( j \) by \( A_j \) and \( B_j \). This alternative procedure would involve "level 2 attribution", with perceptions of intentionality but not self-interest (Blount (1995, p.113)). Our treatment \( C \) procedure is "level 3 attribution", which removes perceptions of both intentionality and self-interest. This provides the comparison we want with treatment \( A \), which is "level 1 attribution" involving perceptions of both intentionality and self-interest. Thus, comparison of data from treatment \( A \) with data from treatment \( C \) provides a measure of the incremental effect of direct reciprocity on subjects' decisions that is not confounded by the possible effect of indirect reciprocity.

All of the above design features were common information given to the subjects except for the aforementioned withholding of the source of the \( A_j \) and \( B_j \) figures in treatment \( C \). The instructions given to the subjects are contained in Appendix \( C \). Each treatment was run in four sessions. There were never fewer than 12 nor more than 18 subjects in a session. The experimental treatments were implemented "across-subjects"; that is, different subjects participated in each of the three treatments.

### 6.4 Subjects’ behavior in the experiment

The subjects' behavior in treatment \( A \), the moonlighting game, is presented in Figure 6.1. The reported figures include the multiplication by three for positive amounts sent by first movers and for negative amounts taken by second movers. We observe that 12 out of the 30 first movers took the maximum of five euros and one subject took one euro from the paired second mover. Three subjects "sent" zero and 14 subjects gave positive amounts

---

\(^2\)See Dufwenberg, Gneezy, Güth, and van Damme (2001) for tests of both direct and indirect reciprocity in the context of the investment game.
to the second mover; five subjects gave the maximum of 30 euros. On average, it was profitable for the first movers to give money to second movers. First movers who sent positive amounts of money to second movers made an average profit of 1.93 euros after the second movers made their return decisions. In contrast, first movers who took money from second movers made an average profit of only 0.15 euros after the second movers made their decisions. Next consider the behavior of second movers. Note that 13 of the 30 second movers neither gave nor took money from first movers. But 17 second movers did reduce their own money payoffs in order to either give or take money from first movers; five of them took money from first movers and 12 gave money to first movers.
6.4.1 Behavior in the moonlighting game vs. behavior in the investment game

The first-mover data from the (moonlighting game) treatment A can be compared with first-mover data from the "no history" investment game treatment in Berg, Dickhaut, and McCabe (1995). In their no history treatment, 30 out of 32 subjects gave positive amounts of money to the second movers and the mean amount given was $15.48 (or the mean amount sent, before multiplication by three, was $5.16). These data appear quite different from the first-mover data from our treatment A. Table 6.1 presents contingency tables comparing first- and second-mover data from their experiment with data from our moonlighting game treatment. Chi-square contingency table tests are used because the two experiments involve different strategy sets.

The contingency table in Table 6.1 for amounts sent by first-movers includes two classes: Class 1 is the number of observations for which the amount sent was nonpositive and Class 2 is the number of observations for which the amount sent was positive. The test reveals a significant difference ($p < .001$) between first-mover data from the two experiments. Thus the extension of the investment game's feasible choice set to include opportunities for agents to take as well as give money has a very significant effect on first movers' decisions.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Send Data Contingency Table</th>
<th>Return Data Contingency Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Class 1 16.71 Class 2 14.29</td>
<td>Class 1 10.76 Class 2 4.36</td>
</tr>
<tr>
<td>Berg et al.</td>
<td>2.29 30.22.71</td>
<td>14.16.36 16.13.64</td>
</tr>
<tr>
<td>Chi-Square Test</td>
<td>16.66 ($p &lt; .001$)</td>
<td>2.36 ($p &gt; .1$)</td>
</tr>
</tbody>
</table>

Table 6.1: Treatment A data vs. Berg et al. data.
CHAPTER 6. TRUST, FEAR, RECIPROCITY, AND ALTRUISM

Next consider amounts returned. Let $S$ denote the amount sent (measured by the cost to the first mover) and $R$ denote the amount returned (measured by the cost to the second mover). The contingency table in Table 6.1 for amounts returned by second movers includes two classes: Class 1 is the number of observations for which $S > 0$ and $R \geq S$; and Class 2 is the number of observations for which $S > 0$ and $R < S$. Class 1 includes subject pairs in which the first mover gave money and the second mover shared the profit on the first mover's "investment", leaving both subjects with a profit. Class 2 includes subject pairs in which the first mover gave money and the second mover defected, leaving the first mover with a loss. The test does not reveal a significant difference between return behavior in the two experiments ($p > .1$). Thus the extension of the feasible choice set, to include the opportunity for both first and second movers to take money, does not significantly change the behavior of second movers who receive positive transfers.

6.4.2 Behavior in our moonlighting game vs. the Abbink et al. moonlighting game

First- and second-mover data from our (moonlighting game) treatment A can be compared with data from the "without contracts" moonlighting game treatment reported in Abbink, Irlenbusch, and Renner (2000). In their experiment, each subject was given a credit balance of 12 "talers", the fictitious currency of the experiment. First movers could send and second movers return integer amounts. Positive amounts sent by first movers were tripled by the experimenters, while negative amounts "sent" were not transformed. A first mover could send an amount that varied from -6 to +6 (or from -6 to +18, including the tripling of positive amounts). Negative amounts "returned" were tripled by the experimenters, while positive amounts returned were not transformed. A second mover could return an amount that varied from -6 to +18 (or from -18 to +18, including the tripling of negative amounts).

In the Abbink et al. (2000) experiment, two out of the 32 first movers took the maximum of six talers and four others took smaller amounts from the paired second mover. Five subjects "sent" zero and 21 subjects gave positive amounts to the second
mover; ten subjects gave the maximum of 18 talers. On average, it was not profitable for the first movers to either give or take money. But the average loss from taking was larger than the average loss from giving money. First movers who gave positive amounts of money to second movers made an average loss of 0.3 talers. First movers who took money from second movers made an average loss of 4.2 talers (or, alternatively, an average loss of 7.4 talers if one excludes the observation excluded by Abbink et al. (2000)).

Next consider the behavior of second movers. In their experiment, 12 of the 32 second movers neither gave nor took money from first movers. But 20 second movers did reduce their own money payoffs in order to either give or take money from first movers; six of them took money from first movers and 14 gave money to second movers.

Table 6.2 presents contingency tables comparing first- and second-mover data from their without-contracts treatment with data from our treatment A. Chi-square contingency table tests are used because the two experiments involve different strategy sets. The contingency table for amounts sent by first-movers includes two classes: Class 1 is the number of observations for which the amount sent was nonpositive and Class 2 is the number of observations for which the amount sent was positive. The test reveals no significant difference (p > .1) between first-mover data from the two experiments.

<table>
<thead>
<tr>
<th></th>
<th>Send Data Contingency Table</th>
<th>Return Data Contingency Table</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Class 1</td>
<td>Class 2</td>
</tr>
<tr>
<td></td>
<td>11.3.94</td>
<td>21.07.07</td>
</tr>
<tr>
<td>Chi-Square Test</td>
<td>2.26</td>
<td>(p &gt; .1)</td>
</tr>
</tbody>
</table>

Table 6.2: Treatment A data vs. Abbink et al. data.

As above, let S denote the amount sent and R denote the amount returned (both measured by the cost to the decision-maker). The contingency table for amounts returned
CHAPTER 6. TRUST, FEAR, RECIPROCITY, AND ALTRUISM

by second-movers includes four classes: Class 1 is the number of observations for which $S \leq 0$ and $-3R \leq S$; Class 2 is the number of observations for which $S \leq 0$ and second mover either increases the money payoff of the first mover or decreases it by $3R$, with $-3R > S$; Class 3 is the number of observations for which $S > 0$ and $R \geq S$; and Class 4 is the number of observations for which $S > 0$ and $R < S$. Note that Class 1 includes subject pairs in which the first mover took money and the second mover strongly retaliated, leaving both subjects with a loss. Class 2 includes subject pairs in which the first mover took money and the second mover retaliated weakly or not at all, leaving the first mover with a profit. Class 3 includes subject pairs in which the first mover gave money and the second mover shared the profit on the first mover’s “investment”, leaving both subjects with a profit. Class 4 includes subject pairs in which the first mover gave money and the second mover defected, leaving the first mover with a loss.

The test reported in Table 6.2 reveals a significant difference between second-mover data from the two experiments ($p < .025$). Inspection of the contingency table reveals notable differences in the number of observations in classes 2 and 4. The larger number of observations in Class 2 for our data is consistent with there being less negative reciprocity in our experiment than in theirs. The larger number of observations in Class 4 for their data is consistent with there being less positive reciprocity in their experiment than in ours.\(^3\)

The source of the difference in second-mover behavior is an interesting issue. There were notable differences between the protocols used in the two experiments. As explained above, our experiment was computerized and involved double-blind payoff and group-separation procedures that created high “social distance” between the first- and second-mover groups of subjects and between the subjects and the experimenters. In contrast, the Abbink et al. (2000) experiment was a “mensa” (cafeteria) experiment in which subjects participated in a manual (non-computerized) experiment as they entered the lobbies of two buildings. Abbink et al. (2000) state that their payoff procedure was “double

\(^3\)Excluding the observation excluded by Abbink et al. leads to the following: (i) the entry in the second row and fourth column of Table 6.2 changes from 1 to 0; (ii) the chi-square test statistic changes from 10.14 to 12.78; and (iii) the $p$-value changes from $p < .025$ to $p < .01$. 
6.4. SUBJECTS' BEHAVIOR IN THE EXPERIMENT

blind" because the subjects were instructed to use pseudonyms. But the description of their procedures does not specify whether the subjects were paid face-to-face by the experimenters, nor does it contain other details that would reveal the level of social distance between the subjects and the experimenters in their protocol. Of course, this discussion of experimental protocols should not be misinterpreted as a view that there is only one proper way to conduct a fairness experiment. To the contrary, running experiments with protocols involving different levels of social distance can yield deeper insight into fairness behavior, most especially about the extent to which social norms for reciprocity and other fairness-oriented behavior are internalized.

6.4.3 Tests of the self-regarding preferences model

We first consider subjects' behavior in treatment A. The traditional model of self-regarding preferences has straightforward predictions for this game. Since a second mover would be assumed to want only to maximize her own money payoff, she would be predicted to neither take nor give money (i.e., return zero) to the paired first mover because either action would be costly to her. Knowing this, and assumed only to want to maximize his own money payoff, a first mover would be predicted to take the maximum of five euros (i.e. send minus five) from the paired second mover.

As noted above, 12 out of the 30 first movers took five euros from their paired second movers and 13 out of the 30 second movers neither gave nor took money from first movers. Aggregating over all subjects in treatment A, 35 out of 60 or 58% of the subjects made decisions that are inconsistent with the self-regarding preferences model. Confining our attention to pairs of subjects, we observe that six of the first movers who took the maximum of five euros were not punished by their paired second movers. Thus the behavior of six out of 30 or 20% of the subject pairs is consistent with the subgame perfect equilibrium of the self-regarding preferences model.

The evidence is mixed concerning whether the level of social distance in a protocol is a significant determinant of behavior in fairness experiments. Hoffman et al. (1994) and Cox and Deck (2001) found significant effects from the single-blind/double-blind treatment while Bolton and Zwick (1995), Bolton, Katok, and Zwick (1998), and Johannesson and Persson (2000) did not.
CHAPTER 6. TRUST, FEAR, RECIPROCITY, AND ALTRUISM

The first row of Table 6.3 reports the means and standard deviations of the amounts sent by first movers and returned by second movers in treatment A. The fourth row of the table reports results from a Kolmogorov test of the hypothesis that amounts sent are equal to minus five; the test implies rejection of the hypothesis. The Kolmogorov test reported in the fifth row of Table 6.3 implies rejection of the hypothesis that amounts returned are equal to zero. We conclude that subjects' behavior in treatment A is not consistent with the predictions of the self-regarding preferences model.

We next describe the subjects' behavior in the dictator treatments. The subjects' behavior in treatment B and treatment C is presented in Figure 6.2 and Figure 6.3.

![Figure 6.2: Money sent in treatments A and B.](image)

The self-regarding preferences model predicts that the maximum of five will be taken by "first movers" in treatment B and that zero will be returned by "second movers" in treatment C. In treatment B, 21 out of 27 subjects took the maximum of five euros, four subjects took smaller amounts, and two subjects gave positive amounts. In treatment C, 21 out of 30 subjects chose zero euros, six subjects "returned" positive amounts, and three subjects "returned" negative amounts. The second and third rows of Table 6.3 report the means and standard deviations of the amounts sent and returned in treatments B and
<table>
<thead>
<tr>
<th>Data</th>
<th>Send Mean</th>
<th>Return Mean</th>
<th>Kolmogorov Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tr. A</td>
<td>7.47</td>
<td>2.10</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>[13.88]</td>
<td>[9.02]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>{30}</td>
<td>{30}</td>
<td></td>
</tr>
<tr>
<td>Tr. B</td>
<td>-3.11</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>[6.83]</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td></td>
<td>{27}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tr. C</td>
<td>...</td>
<td>-0.20</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>[2.91]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>{30}</td>
<td></td>
</tr>
<tr>
<td>Tr. A Send</td>
<td>...</td>
<td>...</td>
<td>0.60</td>
</tr>
<tr>
<td>vs. Minus Five</td>
<td></td>
<td></td>
<td>($p &lt; .01$)(^1)</td>
</tr>
<tr>
<td>Tr. A Ret.</td>
<td>...</td>
<td>...</td>
<td>0.40</td>
</tr>
<tr>
<td>vs. Zero</td>
<td>...</td>
<td>...</td>
<td>($p &lt; .01$)</td>
</tr>
<tr>
<td>Tr. B Send</td>
<td>...</td>
<td>...</td>
<td>0.22</td>
</tr>
<tr>
<td>vs. Minus Five</td>
<td></td>
<td></td>
<td>(.05 &lt; $p &lt; .1$)(^1)</td>
</tr>
<tr>
<td>Tr. C Ret.</td>
<td>...</td>
<td>...</td>
<td>0.20</td>
</tr>
<tr>
<td>vs. Zero</td>
<td>...</td>
<td>...</td>
<td>($p = .15$)</td>
</tr>
<tr>
<td>Tr. A Send</td>
<td>...</td>
<td>...</td>
<td>0.43</td>
</tr>
<tr>
<td>vs. Zero</td>
<td>...</td>
<td>...</td>
<td>($p &lt; .005$)(^1)</td>
</tr>
<tr>
<td>Tr. B Send</td>
<td>...</td>
<td>...</td>
<td>0.93</td>
</tr>
<tr>
<td>vs. Zero</td>
<td>...</td>
<td>...</td>
<td>($p &lt; .005$)(^1)</td>
</tr>
</tbody>
</table>

Table 6.3: Tests of predicted distributions. (Standard deviations in brackets. Number of subjects in braces. \(p\)-values in parentheses. \(^1\) denotes a one-tailed test.)
CHAPTER 6. TRUST, FEAR, RECIPROCITY, AND ALTRUISM

Figure 6.3: Money “sent” and returned in treatment C.

C. The Kolmogorov test reported in row six does not imply rejection of the hypothesis that amounts sent in treatment B are equal to minus five at the 5% level of significance. The Kolmogorov test reported in the seventh row of Table 6.3 does not imply rejection of the hypothesis that amounts returned are equal to zero. Thus the tests do not reject the predictions of the self-regarding preferences model with data from treatments B and C. In treatment A, 42% of the subjects made decisions that are consistent with the self-regarding preferences model. In contrast, in treatments B and C, 42 out of 57 or 74% of the subjects made decisions that are consistent with the (self-regarding preferences) model’s predictions.

6.4.4 Tests for trust, fear, altruism, and reciprocity

The implications of the self-regarding preferences model are inconsistent with the data for the majority of subjects in treatment A. Thus the subjects have motivations that are richer and more complicated than simply a desire to maximize their own money payoffs.
in the experiment. We next examine the information about alternative motivations that is provided by the triadic experimental design.

First consider the behavior of first movers. Figure 6.2 presents the amounts sent in treatments $A$ and $B$. The third row of Table 6.4 reports tests comparing first-mover behavior in treatments $A$ and $B$. From all three of the reported tests we conclude that there is a highly-significant difference between first-mover sending behavior in treatments $A$ and $B$. Thus first movers behave quite differently when the second movers have an opportunity to respond than when they do not.

Altruism could motivate sending positive amounts in either treatment $A$ or $B$. In contrast, first movers' trust in the positive reciprocity of second movers could motivate the sending of positive amounts in treatment $A$ but not in treatment $B$. The experimenters triple any positive amounts sent by first movers. This creates a profit that can be shared between first and second movers in treatment $A$ if the second movers do not defect. Furthermore, first movers' fear of the negative reciprocity of second movers could motivate them to avoid taking money in treatment $A$ but not in treatment $B$. Thus, comparison of subjects' behavior in treatments $A$ and $B$ permits one to discriminate among some alternative motivations.

Altruism is the only motive for first movers to send positive amounts in treatment $B$. As noted above, 25 out of 27 subjects took money in treatment $B$. Only two of the subjects exhibited altruistic motivation by sending positive amounts in treatment $B$. The last row of Table 6.3 reports tests of the hypothesis that amounts sent in treatment $B$ are greater than or equal to zero. Not surprisingly, the tests imply rejection of the hypothesis that amounts sent are non-negative. We conclude that subjects' behavior in treatment $B$ is not characterized by significant altruism. Since the treatment $A$ subjects are drawn from the same subject pool as the treatment $B$ subjects, we conclude that their behavior is also not characterized by significant altruism.

Next consider the question of whether the significantly higher amounts sent by first movers in treatment $A$ than in treatment $B$ are motivated by fear of negative reciprocity or trust in positive reciprocity. A first mover might prefer to take money from the paired second mover. If so, he will take money in treatment $B$; but he may not take money in
### Table 6.4: Test for trust, fear, and reciprocity. (Standard deviations in brackets. Number of subjects in braces. p-values in parentheses. 1 denotes a one-tailed test.)

<table>
<thead>
<tr>
<th>Data</th>
<th>Return Mean</th>
<th>Means Test</th>
<th>Smirnov Test</th>
<th>Mann-Whitney Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Send A</td>
<td>Send A</td>
<td>All Send A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(&lt;0)</td>
<td>(&gt;0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tr. A</td>
<td>-4.54</td>
<td>8.71</td>
<td>2.10</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>[6.84]</td>
<td>[6.78]</td>
<td>[9.02]</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>{13}</td>
<td>{14}</td>
<td>{30}</td>
<td></td>
</tr>
<tr>
<td>Tr. C</td>
<td>-1.46</td>
<td>0.93</td>
<td>-0.20</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>[3.48]</td>
<td>[2.16]</td>
<td>[2.90]</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>{13}</td>
<td>{14}</td>
<td>{30}</td>
<td></td>
</tr>
<tr>
<td>Tr. A Send vs. Tr. B Send</td>
<td>...</td>
<td>3.59</td>
<td>0.53</td>
<td>3.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.000)</td>
<td>(p &lt; .005)1</td>
<td>(p &lt; .001)1</td>
</tr>
<tr>
<td>Tr. A Ret. vs. Tr. C Ret. (send A &gt; 0)</td>
<td>...</td>
<td>4.10</td>
<td>0.71</td>
<td>3.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.000)</td>
<td>(p &lt; .005)1</td>
<td>(p &lt; .001)1</td>
</tr>
<tr>
<td>Tr. A Ret. vs. Tr. C Ret. (send A &lt; 0)</td>
<td>...</td>
<td>-1.45</td>
<td>0.31</td>
<td>-1.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.081)</td>
<td>(p &gt; 0.1)1</td>
<td>(.115)1</td>
</tr>
</tbody>
</table>
6.4. SUBJECTS' BEHAVIOR IN THE EXPERIMENT

treatment A if he is afraid of retaliation (i.e., negative reciprocity). A selfish, fearful first mover would send zero in treatment A and take five in treatment B. A selfish, unafraid first mover would take money in both treatments A and B. As seen in Figure 6.2, 3 out of 30 of the subjects chose zero in treatment A, whereas 25 out of 27 of the subjects took money in treatment B. Thus the behavior of only 11% (= ((3/30) ÷ (25/27)) × 100) of the subjects is consistent with fear of negative reciprocity. One also observes from Figure 6.2 that 13 out of 30 of the first movers in treatment A took money from the paired second mover. Thus the behavior of 47% (= ((13/30) ÷ (25/27)) × 100) of the subjects is inconsistent with fear of negative reciprocity.

A selfish first mover might send a non-positive amount in treatment B but send a positive amount in treatment A because of trust that the second mover would share the profit from the tripling of amounts sent. As seen in Figure 6.2, 14 out of 30 first movers sent positive amounts of money to second movers in treatment A. In contrast, only two first movers sent positive amounts of money to second movers in treatment B. We conclude that many first movers exhibited trust in positive reciprocity in the moonlighting game. If we base our figure on the non-rejection of the hypothesis that subjects sent minus 5 in treatment B, we conclude that 47% (= (14/30) × 100) of the subjects in treatment A exhibited trust. Alternatively, if we base our figure on the difference between the fractions of subjects that sent positive amounts in treatment A (14/30) and treatment B (2/27), we conclude that 39% (= (14/30 − 2/27) × 100) of the subjects in treatment A exhibited trust.

We now consider the behavior of second movers. A “second mover” in treatment C has the same strategy set as a second mover in treatment A. The allocated money payoffs of the first and second movers, prior to the second mover’s decision, are the same in treatments A and C. The difference between the treatments is that first movers’ decisions determine these allocations in treatment A but not in treatment C. Thus, second movers can be motivated by reciprocity in treatment A but not in treatment C. Whether or not the behavior of second movers is characterized by reciprocity is revealed by comparing responses in treatments A and C. Figures 6.1 and 6.3 show how second movers responded to amounts they received in treatments A and C. Figure 6.4 presents a direct comparison
of amounts returned in treatments A and C.

Figure 6.4: Money sent in treatment A and returned in treatments A and C.

We first consider responses by second movers who received positive amounts. Fourteen second movers received positive amounts of money sent by the paired first movers in treatment A and provided by the experimenters in treatment C. How did they respond in each of the two treatments? In treatment A, 11 responded by returning positive amounts to first movers and three second movers kept all of the money. In contrast, in treatment C three second movers returned positive amounts to first movers and 11 second movers kept all of the money. Another striking difference between the treatments is for the five second movers in each treatment who received the maximum of 30 euros. In treatment C, all five of such second movers kept all of the money. In contrast, in treatment A all of them returned positive amounts, with the amounts returned varying from a low of 10 euros to a high of 20 euros. Finally, note that the fourth row of Table 6.4 reports tests comparing amounts returned in treatments A and C by second movers who received positive amounts.
All of the tests detect a highly significant difference between the treatments. We conclude that the behavior of subjects in the moonlighting game is characterized by significant positive reciprocity.

Next consider responses by the 13 second movers who “received” negative amounts in both treatments. Of these subjects, 12 had the maximum amount of five euros taken from them and the other subject had one euro taken. How did they respond in each of the two treatments? In treatment $A$, five second movers responded by incurring a cost to take money from the paired first mover, seven responded by choosing zero, and one responded by giving the first mover one euro. The behavior of the five second movers who took money in treatment $A$ could be explained by either negative reciprocity or inequality aversion. In treatment $C$, three second movers responded by incurring a cost to take money from the paired “first mover”, eight responded by choosing zero, and two responded by giving the “first mover” one euro. The behavior of the three second movers who took money in treatment $C$ could be explained by inequality aversion but not by negative reciprocity. Thus, whether or not the behavior of second movers is characterized by negative reciprocity is revealed by comparing responses in treatments $A$ and $C$. The last row of Table 6.4 reports tests comparing amounts returned in treatments $A$ and $C$ by second movers who received negative amounts. The tests do not detect a significant difference. We conclude that the behavior of subjects in the moonlighting game is not characterized by significant negative reciprocity.

6.5 Concluding remarks

In this chapter we have reported an experiment with a game triad that includes the moonlighting game. Abbink, Irlenbusch, and Renner (2000) had previously reported data for the moonlighting game that are consistent with reciprocity but inconsistent with the traditional self-regarding preferences model. However, their design cannot discriminate among the observable implications of alternative possible motivations. The game triad experiment reported here makes it possible to discriminate among the observable implications for subjects’ choices of trust, fear, reciprocity, and altruism in the moonlighting
Results from our experiment support the conclusion that: 39% to 47% of first movers in the moonlighting game were motivated by trust in the positive reciprocity of second movers; the behavior of second movers who received positive amounts from first movers was characterized by significant positive reciprocity; the behavior of 47% of the first movers is inconsistent with fear of negative reciprocity; and the behavior of second movers who had money taken from them by first movers was not characterized by significant negative reciprocity.

Our experiment supports the conclusion that behavior in the moonlighting game cannot be fully explained by models of preferences over outcomes. Models that include both intentions and preferences over outcomes are needed to fully explain behavior in the moonlighting game.