Available online at www.sciencedirect.com
SCIENCE DIRECT®
Games and Economic Behavior 46 (2004) 260–281
GAMES and Economic Behavior
www.elsevier.com/locate/geb

How to identify trust and reciprocity

James C. Cox

Department of Economics, 401 McClelland Hall, University of Arizona, Tucson, AZ 85721-0108, USA

Received 12 April 2001

Abstract

This paper uses a three-games (or triadic) design to identify trusting and reciprocating behavior. A large literature on single-game trust and reciprocity experiments is based on the implicit assumption that subjects do not have altruistic or inequality-averse other-regarding preferences. Such experimental designs test compound hypotheses that include the hypothesis that other-regarding preferences do not affect behavior. In contrast, experiments with the triadic design do discriminate between transfers resulting from trust or reciprocity and transfers resulting from other-regarding preferences that are not conditional on the behavior of others. Decomposing trust from altruism and reciprocity from altruism or inequality aversion is critical to obtaining empirical information that can guide the process of constructing models that can increase the empirical validity of game theory.
© 2003 Elsevier Inc. All rights reserved.

JEL classification: C70; C91; D63; D64

Keywords: Experimental economics; Game theory; Trust; Reciprocity; Altruism

1. Introduction

In their seminal work on game theory, von Neumann and Morgenstern (1944, 1947) thought it necessary to simultaneously develop a theory of utility and a theory of play for strategic games. In contrast, much subsequent development of game theory has focused on analyzing the play of games to the exclusion of utility theory. In the absence of a focus by game theorists on utility theory, it is understandable that experimentalists testing the theory's predictions have typically assumed that agents' utilities are affine transformations of (only) their own monetary payoffs in the games. This interpretation of game theory incorporates the assumptions that agents do not care about others' (relative or absolute) material payoffs or about their intentions. There is a large experimental literature based on this special-case interpretation of the theory, which I shall subsequently refer to as the model of "self-regarding preferences." The part of the literature concerned with public goods experiments and trust and reciprocity experiments has produced replicable patterns of inconsistency with predictions of the model of self-regarding preferences. For example, the patterns of behavior that have been observed in one-shot trust and reciprocity games are inconsistent with the subgame perfect equilibria of that model. But this does *not* imply that the observed behavior is inconsistent with game theory, which is a point that has not generally been recognized in the literature.

In one prominent research program, the central empirical question has been posed as a contest between game theory and alternative theories based on ideas of cultural or biological evolution.[1] For example, McCabe et al. (1998) pose the question as follows:

> Our objective is to examine game theoretic hypotheses of decision making based on dominance and backward induction in comparison with the culturally or biologically derived hypothesis that reciprocity supports more cooperation than predicted by game theory (p. 10)...

and state their conclusion as

> Contrary to noncooperative game theory, but consistent with the reciprocity hypothesis, many subjects achieve the symmetric joint maximum under the single play anonymous interaction conditions that are expected to give game theory its best shot (p. 22).

Another distinguished research program has focused on inconsistencies between the predictions of principal-agent theory and behavior in experimental labor markets.[2] For example, Fehr et al. (1997, p. 856) conclude that

> Our results indicate, however, that the neglect of reciprocity may render principal agent models seriously incomplete. As a consequence it may limit their predictive power. Moreover, the normative conclusions that follow from models that neglect reciprocity may not be correct.

Widely-disseminated conclusions about robust observations of trust and reciprocity have motivated developments of utility theory intended to improve the empirical validity of game theory. For example, Rabin (1993) and Dufwenberg and Kirchsteiger (2001) have developed models that incorporate perceptions of others' intentions into the utilities of game players. In contrast, Levine (1998), Fehr and Schmidt (1999), and Bolton and Ockenfels (2000) have developed models that incorporate other-regarding preferences (or fairness) into game players' utilities. Models that incorporate both intentions and fairness

E-mail address: jcox@eller.arizona.edu.

0899-8256/$ – see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/S0899-8256(03)00119-2

[1] The research program includes the following papers: Berg et al. (1995), Hoffman et al. (1994, 1996, 1998), McCabe et al. (1996, 1998), and Smith (1998).

[2] The research program includes the following papers: Fehr and Falk (1999), Fehr and Gächter (2000a, 2000b), Fehr et al. (1993, 1996, 1997).

have been developed by Falk and Fischbacher (1999), Charness and Rabin (forthcoming), and Cox and Friedman (2002). But there is a problem with the widely-disseminated conclusions about behavior that are motivating these theory developments: the conclusions are not all supported by the experimental designs that generated the data.

The present paper re-examines some central questions in the literature on trust and reciprocity. It specifically questions the widely-accepted conclusion stated in a recent survey article by Fehr and Gächter (2000b, p. 162):

> Positive reciprocity has been documented in many trust or gift exchange games (for example, Fehr et al., 1993; Berg et al., 1995; McCabe et al., 1996).

The conclusion that positive reciprocity is "documented" by data showing that many proposers send, and responders give back money in trust and gift exchange games is not supported by the experimental designs in the cited papers. The source of the difficulty is that the single-game experimental designs used to generate the data in these experiments do *not* discriminate between actions motivated by trust or reciprocity and actions motivated by other-regarding preferences characterized by altruism or inequality aversion that is not conditional on the behavior of others. In the present paper, a triadic experimental design is used to discriminate between transfers resulting from trust or reciprocity and transfers resulting from other-regarding preferences that are not conditional on the behavior of another. This discrimination is based on dictator games that give a first or "second mover" the same feasible choices as in the original game but eliminate the possible effects of the (observed or anticipated) actions of the other agent. Being able to discriminate between the implications of unconditional other-regarding preferences and trust or reciprocity is important to obtaining the empirical information that can guide the process of formulating a theory of utility that can increase the empirical validity of game theory.

2. Definitions

Interpretations of data in this paper will be based on the following definitions. Preferences over one's own *and* others' material payoffs will be referred to as "other-regarding preferences." Such preferences can be altruistic (Andreoni and Miller, 2002; Cox et al., 2002), inequality-averse (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999), quasi-maximin (Charness and Rabin, forthcoming), or possibly even malevolent. They involve ideas of the fairness of outcomes. Let y^k and y^j denote the money payoffs of agents k and j. Assume that agent k's preferences can be represented by a utility function. Then agent k has other-regarding preferences for the income of agent j if his or her utility function, $u^k(y^k, y^j)$ is *not* a constant function of y^j.

It is important to distinguish between actions motivated by reciprocity and actions motivated by conventional other-regarding preferences that are not conditional on the actions or intentions of others because they have different implications for game-theoretic modeling. The concept of positive reciprocity used in this paper is defined as follows. "Positive reciprocity" is a motivation to repay generous or helpful actions of another by adopting actions that are generous or helpful to the other person. An action that is positively reciprocal is a generous action that is adopted in response to a generous action by another. Thus, positively reciprocal behavior is conditional kindness that is distinct from the unconditional kindness motivated by altruism. An individual who behaves in a reciprocal way makes decisions that can be modeled with other-regarding preferences that are conditional on the perceived intentions behind the actions of others, as in Section 4 and Appendix A.

Suppose that the first mover in an extensive form game chooses an action that benefits the second mover. Further suppose that, subsequently, the second mover adopts an action that benefits the first mover. Is the second mover's action motivated by reciprocity or unconditional other-regarding preferences characterized by altruism or inequality aversion? Section 5 explains how the triadic experimental design discriminates between reciprocity and unconditional other-regarding preferences as explanations for generous second-mover actions.

"Trust" is inherently a matter of the beliefs that one agent has about the behavior of another. An action that is trusting of another is one that creates the possibility of mutual benefit, if the other person is cooperative, and the risk of loss to oneself if the other person defects. If the first mover in an extensive form game believes that the second mover may have other-regarding preferences, or be motivated by positive reciprocity, then the first mover may make an efficiency-increasing transfer to the second mover. The first mover may do this, even if he himself has self-regarding preferences, when he believes that the second mover is unlikely to defect, that is, if he trusts the second mover.

Suppose that the first mover in an extensive form game chooses an action that benefits the second mover. Does the first mover do this because she trusts that the second mover will not defect? Or would she do it anyway because she has other-regarding preferences in which the pair of payoffs created by her action is preferred to the pair of payoffs determined by the two players' endowments? Section 5 explains how the triadic experimental design discriminates between trust and other-regarding preferences as explanations for generous first-mover actions.

The experimental design described in Section 4 involves game triads that include the investment game introduced by Berg et al. (1995) and later used by several other authors.

3. The investment game

The Berg, Dickhaut, and McCabe experimental design for the investment game is as follows. Subjects are divided into two groups, the room A group and the room B group. Each individual subject in each group is given ten $1 bills. Each subject in room B is instructed to keep his or her $10. The subjects in room A are informed that each of them, individually, can transfer to an anonymous paired person in room B any integer number of their own ten $1 bills, from 0 to all 10, and keep the remainder. Any amount transferred by a room A subject is multiplied by 3 by the experimenter before being delivered to a room B subject. Then each room B subject is given the opportunity to return part, all, or none of the tripled amount of the transfer he or she received from the anonymous paired person in room A.

If one assumes that subjects have self-regarding preferences, then game theory predicts that:

(i) room B subjects will keep all of any tripled amounts transferred by room A subjects because room B subjects prefer more money to less; and
(ii) knowing this, room A subjects will not transfer any positive amount.

This subgame perfect equilibrium allocation of the model of self-regarding preferences is Pareto-inferior to some alternative feasible allocations because it leaves each pair of subjects with $20 when it could have ended up with as much as $40.

Results from investment-game experiments reported by Berg, Dickhaut, and McCabe were that the average amount transferred by room A subjects was $5.16 and the average amount returned by room B subjects was $4.66. When data from this experiment were provided to subjects in a subsequent experiment (the "social history" treatment), the average amount transferred by room A subjects was $5.36 and the average amount returned was $6.46. There was large variability across subjects in the amounts transferred and returned. The experiments reported by Berg, Dickhaut, and McCabe used a "double blind" protocol in which subjects' responses were anonymous to other subjects and the experimenters.

Note what is measured by these experiments. A room A subject may be willing to transfer money to a room B person if he trusts that some of the tripled amount transferred will be returned. Further, a room B subject may be willing to return part of the tripled amount transferred if she is motivated by positive reciprocity. But a room A subject may be willing to make a transfer to a paired subject in room B even if there is no opportunity for the latter to return anything. The Berg, Dickhaut, and McCabe experimental design does not allow one to distinguish between transfers resulting from trust and transfers resulting from altruistic other-regarding preferences. Similarly, their design does not provide data that distinguish between second-mover return transfers motivated by reciprocity and returns resulting from unconditional other-regarding preferences. The experimental design used in the present paper makes it possible to discriminate among transfers motivated by trust, reciprocity, and unconditional other-regarding preferences.

4. Experimental design and procedures

The experiment involves three treatments implemented in an across-subjects design. Treatment A is the investment game. Each individual in the second-mover group is credited with a $10 endowment. Each individual in the first-mover group is credited with a $10 endowment and given the task of deciding whether she wants to transfer to a paired individual in the other group none, some, or all of her $10. Any amounts transferred are tripled by the experimenter. Then each individual in the second-mover group is given the task of deciding whether he wants to return some, all, or none of the tripled number of certificates he received to the paired individual in the other group.

Treatment B is a dictator game that differs from treatment A only in that the individuals in the "second-mover" group do not have a decision to make; thus they do not have an opportunity to return any tokens that they receive.

Treatment C involves a decision task that differs from treatment A as follows. First, the "first movers" do not have a decision to make. Each "second mover" is given a $10 endowment. "First movers" are given endowments in amounts equal to the amounts kept (i.e., *not* sent) by the first movers in treatment A. Furthermore, the "second movers" in treatment C are given additional dollar amounts equal to the amounts received by second movers in treatment A from the tripled amounts sent by the first movers in treatment A. The subjects are informed with a table of the exact inverse relation between the number of additional dollars received by a "second mover" and the endowment of the anonymously-paired "first mover."

The experiment sessions are run manually (i.e., not with computers). The payoff procedure is double blind:

(i) subject responses are identified only by letters that are private information of the subjects; and
(ii) monetary payoffs are collected in private from sealed envelopes contained in lettered mailboxes.

Double blind payoffs are implemented by having each subject draw a sealed envelope containing a lettered key from a box containing many envelopes. At the end of the experiment, the subjects use their keys to open lettered mailboxes that contain their monetary payoffs in sealed envelopes. The experimenter is not present in the mailbox room when the subjects collect their payoff envelopes. There is no interaction between the experimenter and the subjects during decision-making parts of an experiment session. All distribution and collection of envelopes containing subject response forms is done by a "monitor" who is randomly selected from the subject pool in the presence of all of the subjects.

All of the above design features are common information given to the subjects except for one item. The subjects in treatment C are *not* informed that the inversely-related amounts of the endowment of the "first mover" and additional certificates of the "second mover" are determined by subjects' decisions in treatment A.[3] The subject instructions and response forms do *not* use the terms "first mover" and "second mover" to refer to the two groups of subjects; instead, the terms "group X" and "group Y" are used. The subjects are assigned randomly to group X and group Y. There were six experiment sessions, two per treatment. No subject participated in more than one experiment session. There were 30 pairs of subjects in treatment B and 32 pairs of subjects in each of treatments A and C.

All of the experiment sessions end with each subject being paid an additional $5 for filling out a questionnaire. First movers and second movers have distinct questionnaires. The questions asked have three functions:

[3] This procedure is followed in order to avoid any possible suggestion of indirect reciprocity (Dufwenberg et al., 2001) to the second movers, which would consist of repaying "first mover" C_j, in treatment C, for the friendly action of first mover A_j, in treatment A.

(i) to provide additional data;
(ii) to provide a check for possible subject confusion about the decision tasks; and
(iii) to provide checks for possible recording errors by the experimenters and counting errors by the subjects.

Subjects do *not* write their names on the questionnaires. The additional data provided by the questionnaires include the subjects' reports of their payoff key letters. Data error checks provided by the questionnaires come from asking the subjects to report the numbers of dollar certificates transferred, received, and returned. These reports, together with two distinct records kept by the experimenters, provide accuracy checks on data recording.

Subjects were recruited with a standardized e-mail message from a computerized database of students that had volunteered to participate in experiments by registering on the web site of the Economic Science Laboratory at the University of Arizona. Some of the subjects had participated in previous economics experiments. The computerized database records the types of experiments that subjects participate in. This information was used to filter subjects that had previously participated in experiments similar to ones reported here from the recruitment e-mail list. Except for this filter, subjects were randomly selected from the database. At the beginning of an experiment session, the subjects were required to show student photo identification cards, print their names on a sign-in form, and write their signatures on the form. Inspection of the sign-in forms verifies that there was no repeat participation.

5. Discriminating between other-regarding preferences and trust or reciprocity

Treatment B differs from treatment A only in that the "second movers" do not have a decision to make; thus they do not have an opportunity to return any part of the tripled amounts sent to them. Since "second movers" cannot return anything in treatment B, first movers cannot be motivated by trust that they will do so. In contrast, in treatment A the first movers may be motivated to send positive amounts by both trust and altruistic other-regarding preferences. Thus conclusions about whether first-mover transfers in the investment game (treatment A) are partially motivated by trust are based on the difference between treatments A and B in the amounts of money sent by first movers to second movers.

Since "first movers" cannot send anything in treatment C, "second movers" cannot be motivated by positive reciprocity, that is, a need to repay a friendly action by a first mover. In contrast, in treatment A, second movers can be motivated to return positive amounts by reciprocity or by unconditional other-regarding preferences. Thus conclusions about whether second-mover transfers in the investment game are partially motivated by reciprocity are based on the difference between treatments A and C in the amounts of money returned by second movers to first movers.

As with any data, one needs a maintained theoretical model to interpret the data from the investment game triadic experiment. I begin by discussing the implications of a model of preferences over outcomes that can be conditional on the behavior of another. This model provides clear testable hypotheses about trust and reciprocity. Subsequently, I discuss some questions that have been raised about this approach.

5.1. Implications of a model of preferences over outcomes

Note that the definition of reciprocity in Section 2 incorporates a possible dependence of preferences over outcomes upon the process that generated those outcomes and beliefs about the behavior of others. Such dependence can provide an explanation of why rational agents undertake actions involving trust and reciprocity. Thus, a first mover can rationally undertake a trusting action if she believes that this choice may trigger a social norm in the second mover that causes him not to defect. Alternatively, a first mover can rationally undertake a trusting action if he believes that the second mover has altruistic or inequality-averse unconditional other-regarding preferences. The experimental design for game triads explained in Section 4 makes it possible to discriminate between the implications of unconditional other-regarding preferences and trust or reciprocity.

I will use the following specific criteria for deciding whether a first mover's behavior is trusting. A first mover will be said to undertake an action in the investment game that exhibits trust if the chosen action:

(i) gives a positive amount of the first mover's money endowment to the second mover; and
(ii) is risky for the first mover, in the sense that the amount of money that is sent is larger than the amount that would maximize the first mover's utility if none were to be returned by the second mover.

Thus a trusting action requires a belief by the first mover that the second mover will not defect and keep too much of the profit generated by the first mover's decision to send a positive amount. If a first mover has self-regarding preferences then the act of sending any positive amount implies trust because such a first mover will lose utility if the second mover does not return at least as much money as the first mover gave up. But a first mover may have other-regarding preferences. Since, in the investment game any amount sent by the first mover is tripled, a first mover with altruistic preferences might prefer to give the second mover some money even if she knew that she would get nothing back. Thus the mere act of sending a positive amount of money is not evidence of trusting behavior unless it is known that first movers have self-regarding preferences. But the treatment B dictator game, together with the treatment A investment game, permit one to identify trusting actions, as follows.

Assume that each subject in every pair has preferences over her own and the paired subject's money payoffs that can be represented by a utility function. These preferences can be other-regarding or self-regarding. If the preferences are self-regarding then the utility function is a constant function of the other's money payoff. If the preferences are other-regarding then they can be altruistic or inequality-averse. In treatment B, a first mover chooses an amount to send from the set, S of integers weakly between 0 and 10. The choice in treatment B, s_b implies

$$u^1(10 - s_b, 10 + 3s_b) \geqslant u^1(10 - s, 10 + 3s), \quad \text{for all } s \in S. \tag{1}$$

Now assume that the amount of money that the first mover gives to the second mover in treatment A, s_a is larger than the amount given in treatment B. Then we can conclude

that the first mover has exhibited trust because the amount sent in treatment A is too large to be fully explained by other-regarding preferences. Thus, if $s_a > s_b$ then we know that the first mover is exposed to risk from the possibility that the second mover will defect and appropriate too much of the money transfer. Specifically, if the second mover were to return nothing in the event that $s_a > s_b$, then statement (1) and strict quasi-concavity of u^1 imply that the first mover will have lower utility than he could have attained if he had known that the second mover would return nothing:

$$u^1(10 - s_a, 10 + 3s_a) < u^1(10 - s_b, 10 + 3s_b) \tag{2}$$

because $s_a \in S$.

Next consider the question of identifying reciprocal behavior. The preferences over payoff (ordered) pairs can be conditioned on a social norm for reciprocity. For example, if the first mover in the investment game sends the second mover some of her money, the second mover may be motivated by a social norm for reciprocity to repay this generous action with a generous response. Within the context of a model of preferences over material payoffs, a social norm for reciprocity can be introduced with a state variable. Thus, the preferences over payoffs can be conditional on a state variable for reciprocity. This is an appropriate representation because, *if* there is reciprocal behavior, *then* individuals behave as if they are more altruistic towards another person after that person has been kind, generous, or trusting. The empirical question is whether or not second movers in the investment game choose more generous actions, after the first mover has intentionally sent them money, than they would in the absence of the first mover's action but the presence of the same money allocation.

When analyzing data from this experiment, I will use the following specific criteria for deciding whether a subjects' behavior is reciprocal. A second mover will be said to undertake an action that exhibits positive reciprocity if the chosen action:

(i) returns to a generous first mover a positive amount of money; and
(ii) is costly to the second mover, in the sense that the amount returned is larger than the amount that would maximize the second mover's utility in the absence of the generous action by the first mover.

A second mover with self-regarding preferences will not return any money to the first mover. But a second mover with either altruistic or inequality-averse other-regarding preferences may return money to the first mover who, after making a positive transfer to the second mover, now has a lower money endowment than the second mover. Thus the mere fact that the second mover returns money to the first mover is not evidence of positive reciprocity. But the treatment C dictator game, together with the treatment A investment game, permits one to identify reciprocal actions, as follows.

A "second mover" in treatment C is given an endowment that is inversely related to the endowment of the paired subject. The endowments of a pair of subjects in treatment C are determined by a (distinct) first mover's decision in treatment A (but the subjects do not know this). Thus, the endowments of a pair of treatment C subjects are given by $(10 - s_a, 10 + 3s_a)$. In treatment C, a "second mover" chooses an amount to return from the set, $R(s_a)$ that contains the integers weakly between 0 and $3s_a$. The choice in treatment C, r_c implies

$$u^2(10 + 3s_a - r_c, 10 - s_a + r_c) \geqslant u^2(10 + 3s_a - r, 10 - s_a + r), \quad \text{for all } r \in R(s_a). \tag{3}$$

Suppose that the second mover returns to the first mover in the investment game a positive amount of money or, perhaps, even a larger amount than the first mover sent: $r_a \geqslant s_a$. This, in itself, does not support a conclusion that the second mover was motivated by positive reciprocity because the assumed choice could have been motivated by maximization of unconditional altruistic or inequality-averse other-regarding preferences. However, if one observes that $r_a > r_c$ then he can conclude that the second mover was motivated by reciprocity because the amount of money returned is too large to be fully accounted for by unconditional other-regarding preferences. This follows from noting that $r_a > r_c$, statement (3), and strict quasi-concavity of u^2 imply

$$u^2(10 + 3s_a - r_a, 10 - s_a + r_a) < u^2(10 + 3s_a - r_c, 10 - s_a + r_c) \tag{4}$$

because $r_a \in R(s_a)$.

It might, at first, seem inconsistent with utility maximization for a subject to return an amount of money, r_a that satisfies inequality (4). But a social norm for reciprocity can change an agent's preferences over material payoffs. Such a norm can be incorporated into a theory of utility by introducing the possibility that an agent's preferences over outcomes can depend on the observed behavior of another. Specifically, with respect to reciprocity, an agent's preferences over his own and another person's material payoffs can depend on whether the other person intentionally helped him or intentionally hurt him or did neither. Thus, let λ_a be a state variable that depends on the amount of money sent by the first mover to the second mover in treatment A:

$$\lambda_a = f(s_a). \tag{5}$$

The utility to the second mover of the monetary payoffs in the investment game can be conditional on the reciprocity state variable. Thus there need be no inconsistency between inequality (4) and the norm-conditional-preference inequality,

$$u^2_{\lambda_a}(10 + 3s_a - r_a, 10 - s_a + r_a) \geqslant u^2_{\lambda_a}(10 + 3s_a - r, 10 - s_a + r), \quad \text{for all } r \in R(s_a). \tag{6}$$

Furthermore, experiments on reciprocal behavior can be characterized as research on the comparative properties of norm-unconditional (u^2), and norm-conditional ($u^2_{\lambda_2}$) utility-maximizing behavior.

A complete model for interpreting data from the triadic investment game experiment is presented in the appendix. Theoretical models that incorporate other-regarding preferences over outcomes that can be conditional on the perceived intentions of others are reported in Falk and Fischbacher (1999), Charness and Rabin (forthcoming), and Cox and Friedman (2002).

In order to incorporate into game theory the possibility that agents can be motivated by reciprocity, one needs to include the possibility that agents' preferences over outcomes may

be conditional on the *observed* behavior of others. But if agents' outcome preferences can be conditional on observations of behavior, can they also be conditional on *anticipations* of behavior?

5.2. *What if outcome preferences can be conditional on anticipated behavior or are not a characteristic of an agent?*

It is conceivable that subjects' outcome preferences could be conditional on anticipations of behavior of others, as illustrated by the following example constructed by a referee. Suppose that the first mover in treatment B gives the paired subject $5, knowing that there is no opportunity for the paired subject to return anything. Also suppose that the first mover in treatment A gives the second mover $5, knowing that the second mover will have an opportunity to share the profit, from the tripling of amounts sent, by returning some money. The zero return in treatment B is determined by the structure of the game. In contrast, if the second mover in the investment game returns zero then the first mover may feel angry and betrayed in addition to not realizing his intended distribution of payoffs. Anticipation of this bad emotional outcome could cause a first mover in the investment game to send less than in the dictator game. If subjects' behavior were consistent with this example, then the test for trusting behavior with data from the triadic design would be a conservative test because a first mover would require an even stronger belief that the second mover would not defect in order to overcome the risks of both sub-optimal money payoffs and bad emotional outcomes. As it turned out, the tests reported in Section 6 do reveal significant trusting behavior. Thus it would not be a problem if the tests were to be conservative, as implied by the preceding example of anticipation-dependent utility of outcomes.

Another referee questioned the central assumption that underlies the triadic experimental design, which is the assumption that preferences are characteristics of agents. The argument was that, while the games in the three treatments may look similar using the author's theoretical framework, we do not know how subjects think about them. It was argued that treatments A, B, and C may elicit different fairness norms, leading to the use of different rules of thumb. The alternative approach advocated by the referee was to use data from experiments with games like treatments A, B, and C to construct a portfolio of rules of thumb that are shortcuts for making decisions in families of situations.

In the following section, I will analyze data from the three treatments using the theoretical framework developed in Section 5.1 and Appendix A. Authors of subsequent papers may want to investigate whether preferences are characteristics of agents in fairness games.

6. Subjects' behavior in the three games

The experiment sessions were conducted in the Economic Science Laboratory at the University of Arizona in November 2000. Similar experiments comparing group and individual behavior in the investment game were conducted in the spring of 1999 and reported in Cox (2002).[4] Subjects' behavior in the investment game will first be discussed. Subsequently, data from all three treatments will be used to ascertain whether there is

[4] Individual subject data from the triadic designs used in both experiments are compared in Cox (2000).

empirical support for the conclusion that the subjects' behavior is characterized by trust and/or reciprocity.

6.1. *First- and second-mover decisions in the investment game*

Figure 1 shows amounts sent and returned by subjects in treatment A, the investment game. There are 32 pairs of subjects. The solid black bar for each numbered subject pair shows the amount sent by the first mover, which will be multiplied by three by the experimenter. The patterned bar for a subject pair shows the amount returned by the second mover. There are six subject pairs, numbered 1–6, for which the first mover sent zero and the second mover returned zero. The behavior of these six pairs is consistent with the subgame perfect equilibrium of the traditional self-regarding preferences model, whereas the behavior of the other 26 subject pairs is inconsistent with that equilibrium. But the consistency of behavior of these six subject pairs must be related to the features of the investment game, as it was implemented by Berg et al. (1995) and in the experiment reported here. If a first mover sends zero then the second mover must return zero. Hence, in this game, *subject-pair* consistency with the above subgame equilibrium prediction is equivalent to consistency of data for only the first-mover. There are nine second movers who received positive transfers but returned zero. The behavior of these nine second movers is consistent with the self-regarding preferences model and it is not constrained to be consistent by the structure of the game.

The first movers in the seven subject pairs numbered 11 to 17 sent exactly one-half of their $10 endowments to the paired second mover. Two of the second movers who received $15, from the $5 amounts sent, kept all of the money. Four of the second movers who were sent $5 returned more than they were sent. And the remaining subject returned $3 to the first mover who sent her or him $5.

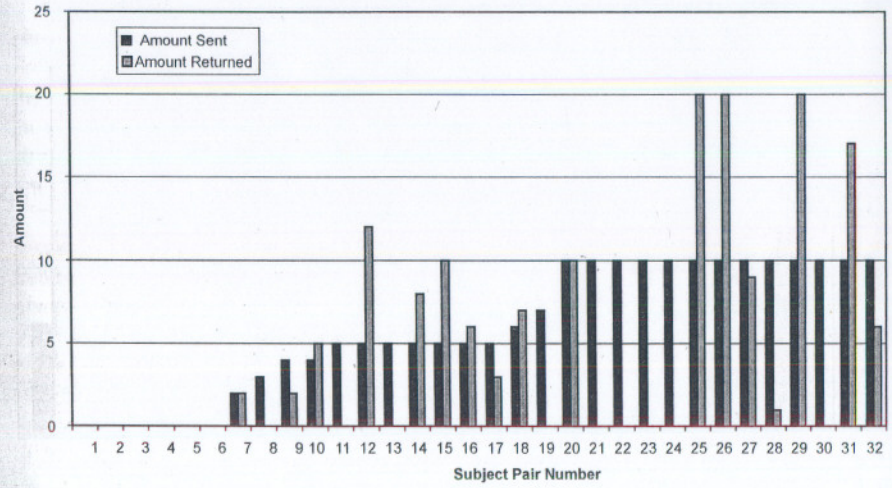

Fig. 1.

The first movers in subject pairs 18 and 19 sent amounts greater than \$5 and less than \$10. One of the paired second movers returned more than was sent and the other second mover returned nothing.

The first movers in the 13 subject pairs numbered 20–32 sent all \$10 of their endowments. The paired second movers exhibited considerable variability in their responses. One of these second movers returned exactly \$10, thus keeping all of the profit from the tripling of the amount sent. Four of the second movers returned nothing, thus ending up with \$40 and leaving their paired first movers with \$0. At the opposite extreme of the data, three of the second movers who received \$30 transfers returned \$20, thus choosing to implement the equal-split fairness focal point payoffs of \$20 for each member of the subject pair. One of the other second movers who was sent \$10 shared the profit by returning \$17. Three other second movers did not share the profit but returned positive amounts of \$1, \$6, and \$9.

As shown in Fig. 1, 26 out of 32 first movers sent positive amounts. Is this trusting behavior? Comparison of behavior in treatments A and B will make it possible to answer this question. Figure 1 also shows that 17 of the second movers returned positive amounts and there appears to be an overall increasing relationship between amounts returned and amounts sent. Is this reciprocal behavior? Comparison of behavior in treatments A and C will make it possible to answer this question.

6.2. *Identifying trust, reciprocity, and altruism*

Figure 2 shows the numbers of first movers in treatments A and B that sent amounts varying from \$0 to \$10. The patterned bars represent treatment A (investment game) data and the solid black bars represent treatment B (trust-control dictator game) data. The first thing to note in Fig. 2 is that 19 out of the 30 first movers in treatment B sent positive amounts of money to the paired subjects. Thus, there is substantial evidence of

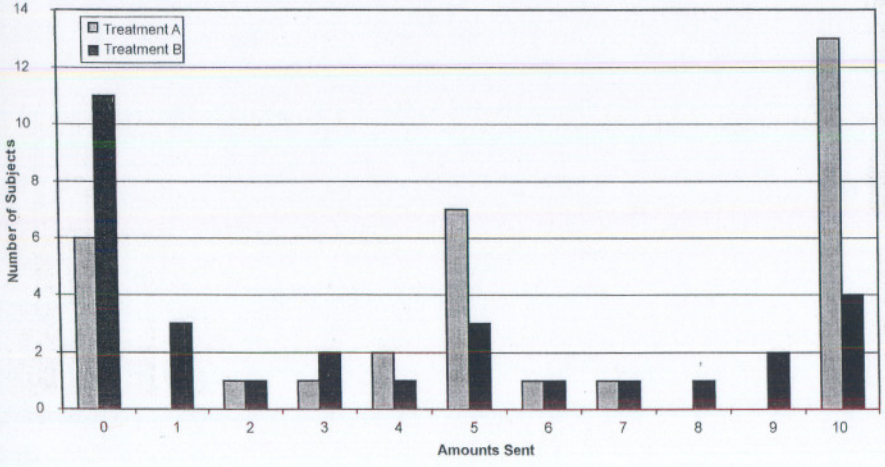

Fig. 2.

Table 1
Decomposition tests for trust and reciprocity

Parametric and nonparametric tests of first- and second-mover data					
Data	Send mean	Return mean	Means tests	Epps–Singleton tests	Mann–Whitney tests
Tr. A	5.97 [3.87] {32}	4.94 [6.63] {32}	…	…	
Tr. B	3.63 [3.86] {30}	…	…	…	
Tr. C	…	2.06 [3.69] {32}	…	…	
Tr. A send vs. Tr. B send	…	…	2.34 (0.010)[a]	16.05 (0.010)	−2.35 (0.010)[a]
Tr. A return vs. Tr. C Return		…	2.88 (0.018)[a]	6.94 (0.219)	−1.55 (0.061)[a]
Tobit analysis of second-mover data					
$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$	LR test	
4.20	0.680	−0.759	0.158	5.98	
(0.060)	(0.034)[a]	(0.124)	(0.008)	(<0.025)	

[a] Denotes a one-tailed test. p-values in parentheses. Standard deviations in brackets. Number of observations in braces.

unconditional other-regarding preferences in these data: when the cost of each dollar sent to the paired subject was only \$0.33, 63% of the subjects behaved as altruists.

Figure 2 shows that six subjects sent \$0 in treatment A whereas 11 subjects made this choice in treatment B. At the other extreme, 13 subjects sent all \$10 in treatment A whereas four subjects made this decision in treatment B. This pronounced difference suggests that the first movers' behavior in treatment A partly resulted from trust. Another notable difference in Fig. 2 is at \$5: seven first movers sent that amount in treatment A but only three did so in treatment B. Finally, note that there is more variability of behavior in treatment B data, with six subjects sending amounts of \$1, \$8, or \$9 that are not observed in treatment A.

Is there statistically-significant support for the existence of trust in the data? The second column of Table 1 reports that the mean amount sent by first movers was \$5.97 in treatment A and \$3.63 in treatment B. The mean amount sent in treatment A is significantly greater than that in treatment B by the one-tailed two-sample t-test ($p = 0.010$) reported in the fourth column of Table 1. Hence the means test supports the conclusion that the subjects exhibited trust in the investment game. As reported in Table 1, the one-tailed Mann–Whitney test also detects that the treatment A amounts sent are significantly greater than the treatment B amounts sent ($p = 0.010$). The Epps–Singleton test detects a significant difference between the cumulative distributions of amounts sent in treatments A and B ($p = 0.010$). Hence all of these tests support the conclusion that there is significant trusting behavior in the investment game.

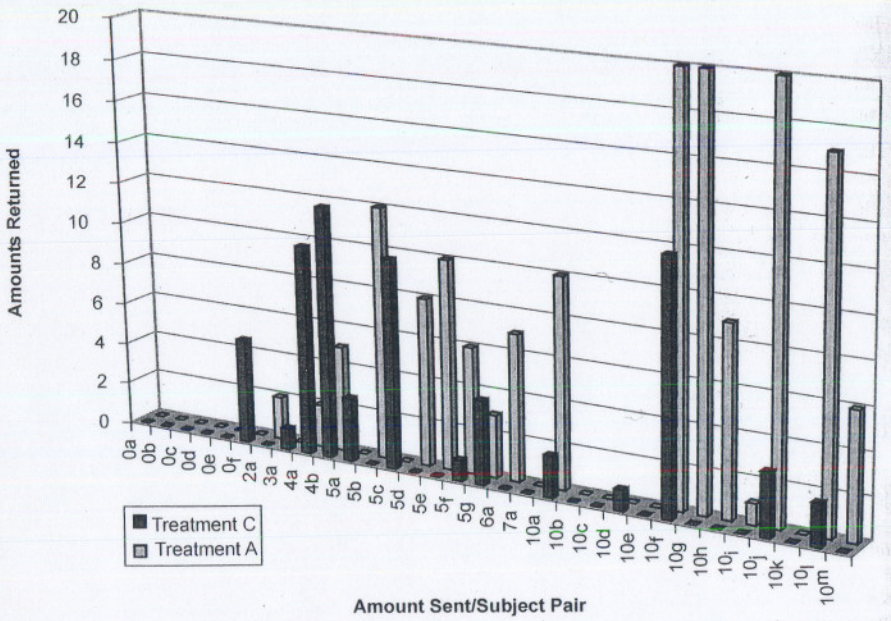

Fig. 3.

In Fig. 3, the patterned bars show the amounts returned in treatment A (the investment game) and the solid black bars show the amounts returned in treatment C (the reciprocity-control dictator game). The first thing to note in Fig. 3 is that 13 out of the 32 "second movers" in treatment C "returned" positive amounts of money to the paired subjects. Thus, there is substantial evidence of unconditional other-regarding preferences in these data: when the cost of each dollar sent to the paired subject was as high as \$1, 41% of the subjects behaved as though they had altruistic or inequality-averse other-regarding preferences.

The floor axis in Fig. 3 records the amounts sent by first movers. The floor axis is labeled with number/letter combinations. The number is the amount sent and the letter designates a first mover who sent that amount in treatment A. Some notable differences between treatments A and C show up in Fig. 3. First consider the 13 observations for which the amount sent was \$10. For this category, five out of the 13 second movers in treatment A returned amounts greater than or equal to \$10. In contrast, only one out of the 13 "second movers" in treatment C that were "sent" \$10 returned an amount greater than or equal to \$10. Another notable difference appears with the nine observations for which the amount sent varied from \$5 to \$7. For this category, five out of the nine second movers in treatment A returned more than was sent. In contrast, only one out of the nine "second movers" in treatment C that were "sent" amounts between \$5 and \$7 "returned" an amount greater than or equal to the amount "sent." There are three observations for which the amounts "returned" in treatment C exceed the amounts returned in treatment A when the amounts sent are low, varying from \$0 to \$4.

Is there statistically-significant support for the existence of reciprocity in the data? The third column of Table 1 reports that the mean amount returned by second movers was \$4.94 in treatment A and \$2.06 in treatment C. The mean amount returned in treatment A is significantly greater than that in treatment C by the one-tailed two-sample t-test ($p = 0.018$) reported in the fourth column of Table 1. The one-tailed Mann–Whitney test also detects that the treatment A amounts returned are significantly greater than the treatment B amounts returned ($p = 0.061$). The Epps–Singleton test does not detect a significant difference between the cumulative distributions of amounts returned in treatments A and C ($p = 0.219$).

The last row of Table 1 reports tobit estimates of the parameters of the following relation between amounts sent, S_t and amounts returned, R_t in treatments A and C:

$$R_t = \alpha + \beta D_t S_t + \gamma S_t + \varepsilon_t, \tag{7}$$

where

$$D_t = \begin{cases} 1 & \text{for treatment A data,} \\ 0 & \text{for treatment C data.} \end{cases} \tag{8}$$

The bounds for the tobit estimation are the bounds imposed by the experimental design:

$$R_t \in [0, 3S_t]. \tag{9}$$

One would expect that the cone created by these bounds might produce heteroskedastic errors. In order to allow for the possibility of heteroskedastic errors, the tobit estimation procedure incorporates estimation of the θ parameter in the following model of multiplicative heteroskedasticity:

$$\sigma_t = \sigma e^{\theta S_t}. \tag{10}$$

Note that $\hat{\beta}$ is the estimate of the effect of reciprocity on amounts returned by second movers. We observe that $\hat{\beta}$ is positive and significantly greater than 0 ($p = 0.034$); hence the tobit estimation supports the conclusion that the subjects exhibited positive reciprocity in the investment game. As noted above, the means test and Mann–Whitney test support the same conclusion.

7. Concluding remarks

This paper reports experiments with a triadic design that can identify trusting and reciprocating behavior. Several researchers had previously reported the replicable result that the majority of first movers send positive amounts and the majority of second movers return positive amounts in investment game experiments. This pattern of results, and results from many other fairness experiments, are inconsistent with the subgame perfect equilibria of the special case of game theory in which players are assumed to have self-regarding preferences. This leaves the profession with the task of constructing a less restrictive model that can maintain consistency with the empirical evidence. But this task cannot be undertaken successfully unless we can discriminate among the observable implications of alternative causes of the deviations from behavior predicted by the self-regarding preferences model. The game triad experiments reported here make it possible to

discriminate among the observable implications for subjects' choices of trust, reciprocity, and unconditional other-regarding preferences. This discrimination is possible because:

(i) treatments A and B jointly identify the trusting behavior that results from beliefs about others; and
(ii) treatments A and C jointly identify the reciprocating behavior that results from imputations of the intentions of others.

There are a few other studies that have used control treatments for intentions. Blount (1995) compared second mover rejections in a standard ultimatum game with second mover rejections in games in which the first move was selected randomly or by an outside party rather than by the subject that would receive the first mover's monetary payoff. She found lower rejection rates in the random treatment than in the standard ultimatum game and lower or similar rejection rates in the third party and standard games, depending upon the choice of elicitation mode for subjects' decisions. Charness (forthcoming) used Blount's control treatments in experiments with the gift exchange game. He found somewhat *higher* average second mover contributions in the outside party and random treatments than in the standard gift exchange game. The average figures reported by Charness reflect lower second mover contributions in the gift exchange game than in the control treatments at low wage rates, a result that is consistent with negative reciprocity. Bolton et al. (1998) experimented with an intentions-control treatment in the context of simple dilemma games. In the control treatment, the row player "chooses" between two identical rows of monetary payoffs. They found no significant differences between the column players' responses in the control treatments and the positive and negative reciprocity treatments.

Our experiment provides evidence of altruistic other-regarding preferences, trust, and reciprocity. These results have the following implications for constructing a model that will be consistent with the observed behavior. First, utility should not be assumed to be a constant function of others' money payoffs, as in the self-regarding preferences model. This is required in order to maintain consistency with the treatment B and C dictator games in which the majority of subjects give money to the paired subjects knowing that the paired subjects have no decision to make. Second, beliefs about others' altruistic and reciprocating behavior should be incorporated in the model. This is required in order to maintain consistency with the trusting behavior that is jointly identified by the investment game (treatment A) and the beliefs-control dictator game (treatment B). Third, the other-regarding preferences should be conditional on the perceived intentions behind others' actions. This is required in order to maintain consistency with the reciprocating behavior that is jointly identified by the investment game (treatment A) and the intentions-control dictator game (treatment C).

Acknowledgments

Financial support was provided by the Decision Risk and Management Science Program, National Science Foundation (grant number SES-9818561). I am grateful to a journal referee for helpful comments and suggestions.

Appendix A. Testable hypotheses derived from the triadic experimental design

I shall explain the structure of the three games and model the players' (utility) payoffs in a general way. Each player's utility function will explicitly incorporate the monetary income of the paired player. It is important to understand that I am *not* assuming that the game players necessarily have other-regarding preferences; instead, I am allowing for that possibility. The subjects' behavior in the experiment with the three games informs us as to whether they do or do not have other-regarding preferences. The second mover's utility function will explicitly incorporate a state variable that introduces the possibility that a trusting action by the paired first mover could trigger an internalized social norm that affects the second mover's utility of the two players' money payoffs from the game. It is also important to understand that I am *not* assuming that the game players necessarily are affected by social norms for reciprocity but am, rather, including that as a possibility. Once again, it is the subjects' behavior in the experiment that informs us on this question.

A.1. Treatment A

Treatment A is the investment game, which can be modeled as follows. The first mover chooses $s_a \in S$, where

$$S = \{0, 1, 2, \ldots, 10\}. \tag{A.1}$$

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

$$R(s_a) = \{0, 1, 2, \ldots, 3s_a\}. \tag{A.2}$$

At the time the first mover makes her choice of s_a, she may not know what choice the second mover will subsequently make. Let the random variable $\tilde{r}$ with probability distribution function $\Omega(\tilde{r}|s_a)$, defined on $R(s_a)$, represent the first mover's beliefs about the amount of money that will be returned by the second mover in subgame $\Gamma(s_a)$.

The first mover's expected payoff from choosing s_a in game A is

$$EP_A^1 = \underset{\Omega(\tilde{r}|s_a)}{E}\left[u^1(10 - s_a + \tilde{r}, 10 + 3s_a - \tilde{r})\right]. \tag{A.3}$$

In the special case where the first mover has self-regarding preferences, u^1 is a constant function of the second mover's income.

A.2. Treatment B

Treatment B is a dictator game with the same strategy set for the first mover as in the investment game. Thus the first mover chooses $s_b \in S$, where S is defined in statement (A.1). The "second mover" does not have a decision to make. The (utility) payoff to the first mover is

$$P_B^1 = u^1(10 - s_b, 10 + 3s_b). \tag{A.4}$$

A.3. Treatment C

Treatment C involves a game $C(n)$, that is selected by the choice made by a first mover in treatment A. It is a dictator game with the same strategy set for the "second mover" that a second mover has in treatment A. Thus the "second mover" chooses $r_c \in R(s_a)$. The (utility) payoff to the "second mover" in game $C(n)$ will not be dependent on the possible operation of a social norm for reciprocity because the first mover has no decision to make in this game.

A.4. Payoffs dependent on social norms

The utility to the second mover of the monetary payoffs from a game can be made conditional on the possible operation of a social norm for reciprocity. Thus, the payoff to the second mover from the choices of s_a and r_a in game A will be written as

$$P_A^2 = u_{\lambda_a}^2(10 + 3s_a - r_a, 10 - s_a + r_a) \tag{A.5}$$

because the second mover knows that the first mover has chosen the action s_a and may feel obliged to reciprocate. The notation $u_{\lambda_a}^2$ permits the utility of monetary payoffs to vary with a state variable λ_a that depends on the amount of money sent by the first mover to the second mover in treatment A:

$$\lambda_a = f(s_a). \tag{A.6}$$

In contrast, in game $C(n)$ the "second mover" knows that the "first mover" has no decision to make. Since there is no opportunity for trusting actions by the "first mover" in game $C(n)$, there is no reason for a social norm for reciprocating to be triggered. Thus the payoff to the "second mover" from the choice of r_c in game $C(n)$ will be written as

$$P_{C(n)}^2 = u^2(10 + 3s_a - r_c, 10 - s_a + r_c). \tag{A.7}$$

In the special case where a social norm for reciprocity does not affect utility of monetary payoffs, $u_{\lambda_a}^2$ is identical to u^2 for all $s_a \in S$.

A.5. Testing for the presence of trust

In order validly to conclude that a first mover has demonstrated trust, the researcher must have knowledge that she has borne a risk of loss from her choice in game A. Thus it must be known that there exists $r_z \in R(s_a)$ and $s_\tau \in S$ such that

$$u^1(10 - s_a + r_z, 10 + 3s_a - r_z) < u^1(10 - s_\tau, 10 + 3s_\tau). \tag{A.8}$$

Assuming that u^1 is strictly quasi-concave (and recalling that the variables are discrete), the choices by the first mover allow the researcher to conclude that (A.8) is satisfied by $r_z = 0$ and $s_\tau = s_b$ if

$$s_a > s_b + 1. \tag{A.9}$$

This can be seen by noting that the choice by the first mover in game B and strict quasi-concavity of u^1 imply

$$u^1(10 - s_b, 10 + 3s_b) > u^1(10 - s, 10 + 3s), \quad \forall s \in S,\ s > s_b + 1. \tag{A.10}$$

The null hypothesis is that the self-regarding preferences model makes empirically-correct predictions. In the present context, this means that the first mover has *not* exhibited trust:

$$H_0^{\mathrm{T}}\colon\ s_a \leqslant s_b + 1, \tag{A.11}$$

with alternative

$$H_{\mathrm{A}}^{\mathrm{T}}\colon\ s_a > s_b + 1. \tag{A.12}$$

It may seem unlikely that the first mover will be indifferent between s_b and $s_b + 1$; hence the null hypothesis in statement (A.11) may seem to bias the tests against finding that the data contain evidence of trust. Furthermore, across-subjects comparisons between treatments involve means and other aggregations of data for which the \$1 unit of discreteness does not apply. Therefore, the tests reported are for the null hypothesis,

$$H_{00}^{\mathrm{T}}\colon\ s_a \leqslant s_b, \tag{A.13}$$

with alternative given by

$$H_{\mathrm{AA}}^{\mathrm{T}}\colon\ s_a > s_b. \tag{A.14}$$

Of course, the hypotheses that are tested statistically will be stochastic versions of H_{00}^{T}.

A.6. Testing for the presence of reciprocity

In order validly to conclude that a second mover has demonstrated positive reciprocity, the researcher must have knowledge that the second mover has incurred a cost to repay a social debt to the first mover. This can be manifested by the second mover choosing to return an amount of money in game A that is larger than the amount that would maximize his utility in the absence of a social norm for reciprocating. Thus, the second mover has exhibited positive reciprocity in game A if there exists $r_y \in R(s_a)$ such that

$$u^2(10 + 3s_a - r_y, 10 - s_a + r_y) > u^2(10 + 3s_a - r_a, 10 - s_a + r_a). \tag{A.15}$$

Assuming that u^2 is strictly quasi-concave (and recalling that the variables are discrete), the choices by the second mover allow the researcher to conclude that (A.15) is satisfied if

$$r_a > r_c + 1. \tag{A.16}$$

This can be seen by noting that the choice by the "second mover" in game $C(n)$ and strict quasi-concavity of u^2 imply

$$u^2(10 + 3s_a - r_c, 10 - s_a + r_c) > u^2(10 + 3s_a - r, 10 - s_a + r),$$
$$\forall r \in R(s_a),\ r > r_c + 1. \tag{A.17}$$

The null hypothesis is that the self-regarding preferences model makes empirically-correct predictions. In the present context, this means that the second mover has *not* exhibited reciprocity:

$$H_0^{\mathrm{R}}\colon\ r_a \leqslant r_c + 1, \tag{A.18}$$

with alternative

$$H_{\mathrm{A}}^{\mathrm{R}}\colon\ r_a > r_c + 1. \tag{A.19}$$

For the reasons explained above in the context of testing for trust, the reported tests for reciprocity are based on stochastic versions of

$$H_{00}^{R}\colon\ r_a \leqslant r_c, \tag{A.20}$$

with alternative given by

$$H_{AA}^{R}\colon\ r_a > r_c. \tag{A.21}$$

References

Andreoni, J., Miller, J., 2002. Giving according to GARP: An experimental test of the consistency of preferences for altruism. Econometrica 70, 737–753.

Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games Econ. Behav. 10, 122–142.

Blount, S., 1995. When social outcomes aren't fair: the effect of causal attributions on preferences. Organ. Behav. Human Dec. Processes 63, 131–144.

Bolton, G., Brandts, J., Ockenfels, A., 1998. Measuring motivations for the reciprocal responses observed in simple dilemma games. Exper. Econ. 1, 207–219.

Bolton, G., Ockenfels, A., 2000. ERC: a theory of equity, reciprocity and competition. Amer. Econ. Rev. 90, 166–193.

Charness, G., forthcoming. Attribution and reciprocity in an experimental labor market. J. Lab. Econ.

Charness, G., Rabin, M., forthcoming. Understanding social preferences with simple tests. Quart. J. Econ.

Cox, J., 2000. Trust and reciprocity: implications of game triads and social contexts. Univ. of Arizona discussion paper.

Cox, J., 2002. Trust, reciprocity, and other-regarding preferences: groups vs. individuals and males vs. females. In: Zwick, R., Rapoport, A. (Eds.), Advances in Experimental Business Research. Kluwer.

Cox, J., Friedman, D., 2002. A tractable model of reciprocity and fairness. Univ. of Arizona working paper.

Cox, J., Sadiraj, K., Sadiraj, V., 2002. A theory of competition and fairness for egocentric altruists. Univ. of Arizona working paper.

Dufwenberg, M., Kirchsteiger, G., 2001. A theory of sequential reciprocity. Discussion paper. University of Vienna.

Dufwenberg, M., Gneezy, U., Güth, W., van Damme, E., 2001. Direct versus indirect reciprocity: an experiment. Homo Oeconom. 18, 19–30.

Falk, A., Fischbacher, U., 1999. A theory of reciprocity. Working paper No. 6. Institute for Empirical Research in Economics, University of Zurich.

Fehr, E., Falk, A., 1999. Wage rigidity in a competitive incomplete contract market. J. Polit. Econ. 107, 106–134.

Fehr, E., Gächter, S., 2000a. Do incentive contracts crowd out voluntary cooperation? Univ. of Zurich working paper.

Fehr, E., Gächter, S., 2000b. Fairness and retaliation: the economics of reciprocity. J. Econ. Perspect. 14, 159–181.

Fehr, E., Gächter, S., Kirchsteiger, G., 1996. Reciprocal fairness and noncompensating wage differentials. J. Inst. Theoret. Econ. 152, 608–640.

Fehr, E., Gächter, S., Kirchsteiger, G., 1997. Reciprocity as a contract enforcement device: experimental evidence. Econometrica 65, 833–860.

Fehr, E., Kirchsteiger, G., Riedl, A., 1993. Does fairness prevent market clearing? An experimental investigation. Quart. J. Econ. 108, 437–460.

Fehr, E., Schmidt, K., 1999. A theory of fairness, competition, and cooperation. Quart. J. Econ. 114, 817–868.

Hoffman, E., McCabe, K., Shachat, K., Smith, V., 1994. Preferences, property rights, and anonymity in bargaining games. Games Econ. Behav. 7, 346–380.

Hoffman, E., McCabe, K., Smith, V., 1996. Social distance and other-regarding behavior in dictator games. Amer. Econ. Rev. 86, 653–660.

Hoffman, E., McCabe, K., Smith, V., 1998. Behavioral foundations of reciprocity: experimental economics and evolutionary psychology. Econ. Inquiry 36, 335–352.

Levine, D., 1998. Modeling altruism and spitefulness in experiments. Rev. Econ. Dynam. 1, 593–622.

McCabe, K., Rassenti, S., Smith, V., 1996. Game theory and reciprocity in some extensive form bargaining games. Proc. Nat. Acad. Sci. 93, 13421–13428.

McCabe, K., Rassenti, S., Smith, V., 1998. Reciprocity, trust, and payoff privacy in extensive form bargaining games. Games Econ. Behav. 24, 10–24.

Rabin, M., 1993. Incorporating fairness into game theory and economics. Amer. Econ. Rev. 83, 1281–1302.

Smith, V., 1998. Distinguished guest lecture: the two faces of Adam Smith. Southern Econ. J. 65, 1–19.

von Neumann, J., Morgenstern, O., 1944. Theory of Games and Economic Behavior, 1st edition. Princeton Univ. Press, 2nd edition, 1947.