On the Economics of Reciprocity

James C. Cox
University of Arizona
jcox@bpa.arizona.edu

January 2001

On the Economics of Reciprocity*

1. Introduction

Much of economic and game theoretic modeling has focused on the special case in which agents are assumed to be exclusively concerned with maximizing their own material payoffs. This model of “self-regarding preferences” has sharp empirical implications in a wide variety of contexts. And the model predicts behavior quite well in many types of controlled experiments. For example, bids and offers in double auction markets for items with known values converge to competitive outcomes (Smith, 1982; Davis and Holt, 1993). This double auction convergence is robust to very unequal gains from exchange (Smith and Williams, 1990). Experiments with “proposer competition” (Roth, et al., 1991) and “responder competition” (Güth, Marchand, and Rulliere, 1997) also produce outcomes in which very unequal material payoffs are accepted by the subjects. And the self-regarding preferences model predicts well in a variety of contexts such as one-sided auctions with independent private values (Cox and Oaxaca, 1996), procurement contracting (Cox, et al., 1996), and search (Cox and Oaxaca, 1989, 2000; Cason and Friedman, 2000; Harrison and Morgan, 1990).

Although the self-regarding preferences model predicts quite well in many contexts, there is now a large body of experimental literature that has produced replicable patterns of inconsistencies with the model’s predictions in contexts involving salient fairness considerations or opportunities for cooperation. For example, the model predicts zero contributions in voluntary contributions public goods experiments but many people make positive contributions, most especially when there are costly opportunities for punishing free riders (Fehr and Schmidt, 1999). The self-regarding preferences model predicts that proposers will ask for all or nearly all of the money in ultimatum games and that responders will accept such proposals. But most proposers ask for nearly equal shares and many responders veto proposals in which they would receive less than a 30% share in the total payoff (Güth, Schmittberger, and Schwarze, 1982; Slonim and Roth, 1997). There are systematic differences between the predictions of standard principal-agent theory and behavior in experimental labor markets. In the presence of incomplete contracts, wage offers above the opportunity cost of workers elicit effort choices above the predicted (shirking) level (Fehr, Gächter, and Kirchsteiger, 1997).[1] In a variety of other types of experiments, decision-makers deviate from the predictions of the traditional model in ways that can be attributed to positive and negative reciprocity.

In a recent survey paper, Fehr and Gächter (2000) reviewed a large literature on the "economics of reciprocity." The common feature of the reviewed papers is the authors' conclusion that deviations from the predictions of the self-regarding preferences model are explained by positive or negative reciprocity or related motivations such as trust (in positive reciprocity) or fear (of negative reciprocity).

Fehr and Gächter (2000, p. 159) define reciprocity as follows: "Reciprocity means that in response to friendly actions, people are frequently much nicer and much more cooperative than predicted by the self-interest model; conversely, in response to hostile actions they are frequently much more nasty and even brutal." Fehr and Gächter (2000, p. 160) draw a clear distinction between reciprocity and altruism: "Reciprocity is also fundamentally different from altruism. Altruism is a form of *unconditional* kindness; that is, altruism given does not emerge as a response to altruism received."

The present paper raises questions about the economics of reciprocity literature precisely because most of it reports results from experimental designs that cannot discriminate between choices motivated by reciprocity and choices motivated by altruism. A critique with applicability to a variety of environments is developed within the specific context of the trust (or investment) game described and interpreted by Fehr and Gächter (2000, p. 162) as follows:

> *Positive reciprocity has been documented in many trust or gift exchange games (for example, Fehr, Kirchsteiger, and Riedl, 1993; Berg, Dickhaut, and McCabe, 1995;*

> *McCabe, Rassenti, and Smith, 1996). In a trust game, for example, a Proposer receives an amount of money x from the experimenter, and then can send between zero and x to the Responder. The experimenter then triples the amount sent, which we term y, so the Responder has 3y. The Responder is then free to return anything between zero and 3y to the Proposer. It turns out that many Proposers send money and many Responders give back some money.*

There is a problem with 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. The problem is that the single-game experimental designs used to generate the data in these experiments do *not* discriminate between actions motivated by reciprocity and actions motivated by altruism. It is true that the self-regarding preferences model predicts: (a) responders will keep all of any tripled amounts sent by proposers because the responders prefer more (of their own) money to less; and (b) knowing this, the proposers will not transfer any positive amount because they prefer more (of their own) money to less. But a responder may be willing to return money to the proposer because of altruism rather than reciprocity. And a proposer may be willing to send money to a responder because of altruism rather than any trust that the responder will share part of the profit from the experimenter tripling the amount sent. Therefore, a more elaborate experimental deign is needed to generate data that can document, or fail to document, positive reciprocity in trust or gift exchange games.

Use of experimental designs that can discriminate between actions motivated by altruism or by reciprocity is essential to obtaining data that can inform the efforts to develop models that can increase the empirical validity of game theory. We want to ascertain whether there are contexts in which deviations from the predictions of the self-regarding preferences model can be completely explained by other-regarding preferences (such as altruism) or by internalized norms for responding to others' actions (such as reciprocity). More generally, the research objective is to discover contexts in which one or the other type of motivation is the quantitatively more significant determinant of behavior.

This paper uses an experimental design that decomposes proposer and responder transfers in the investment (or trust) game into the parts, if any, motivated by altruism and the parts, if any, motivated by responders' reciprocity or proposers' trust.

2. Experimental Design and Procedures

In order to be able to discriminate empirically between actions motivated by reciprocity or altruism and between actions motivated by trust or altruism, one needs to run experiments with a triadic experimental design. The three games used here are the investment game, a specific dictator control game for proposer motivations, and another specific dictator control game for responder motivations.

The experiment involves three treatments. Treatment A implements the investment game of Berg, Dickhaut, and McCabe (1995). Each responder is credited with a $10 endowment. Each proposer is credited with a $10 endowment and given the task of deciding whether he wants to transfer ("send") to an anonymously paired responder none, some, or all of his endowment. Any amounts sent are tripled by the experimenter. Then each responder is given the task of deciding whether she wants to return to the anonymously paired proposer some, all, or none of the tripled amount sent to her.

Treatment B differs from treatment A only in that the responders 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 responders cannot return anything in treatment B, proposers cannot be motivated by trust that they will do so. In contrast, in treatment A proposers may be motivated to send positive amounts by trust or by altruistic other-regarding preferences. Thus the measure of transfers motivated by trust is the difference between treatments A and B in the amounts sent by proposers to responders.[2]

Treatment C involves a decision task that differs from Treatment A as follows. First, the proposers do not have a decision to make. Each responder is given a $10 endowment. Proposers are given endowments in amounts equal to the amounts kept (i.e. *not* sent) by the proposers in treatment A. Furthermore, the responders in treatment C are given additional dollars in amounts equal to the amounts received by responders in treatment A from the tripled amounts sent by the proposers in the latter treatment. The subjects are informed with a table of the exact inverse relation between the number of additional dollars received by a responder and the endowment of the anonymously paired proposer. Since proposers cannot send anything in treatment C, responders cannot be motivated by a need to repay a friendly action by a proposer (i.e., positive reciprocity). In contrast, in treatment A responders may be motivated to return positive amounts by positive reciprocity or by altruistic other-regarding preferences. Thus the measure of transfers due to reciprocity is the difference between the amount of money returned in treatments A and C.[3]

The experiment sessions are run manually (i.e., not with computers). The payoff procedure is double blind: (a) subject responses are identified only by letters that are private information of the subjects; and (b) 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 proposer endowments and the responder additional dollars are determined by subject decisions in treatment A.[4] The subject instructions and response forms do *not* use the terms "proposer" and "responder" 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 subjects in treatment B and 32 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. Proposers and responders have distinct questionnaires. The questions asked have three functions: (a) to provide additional data; (b) to provide a check for possible subject confusion about the decision tasks; and (c) 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 tokens transferred, received, and returned. These reports, together with two distinct records kept by the experimenters, provide accuracy checks on data recording.

The experiment was run at the University of Arizona in November 2000. A detailed explanation of the experiment procedures and the subject instructions are contained in an appendix available upon request to the author.

3. Altruism and/or Trust and Reciprocity

The second column of Table 1 reports that the mean amount sent by proposers to responders 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 = .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

Kolmogorov-Smirnov two-sample test also detects that the treatment A amounts sent are significantly greater than the treatment B amounts sent (*p* = .045), as does the Mann-Whitney test (*p* = .010); hence these tests lend further support to the finding that there is significant trusting behavior in the investment game.

Data from treatments A and B can be used to decompose the total amounts sent by proposers into amounts sent because of trust and amounts sent because of altruism. Using the mean amounts sent by proposers in treatment A and treatment B, we conclude that 61% (= 100($3.63/$5.97)) of the mean amount sent by proposers to responders can be explained by altruism. Thus, contrary to findings previously reported in the literature, the quantitatively more important reason that the data are inconsistent with the zero-transfer prediction of the model of self-regarding preferences is altruism, not trust.

The third column of Table 1 reports that the mean amount returned by responders to proposers 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* = .018) reported in the fourth column of Table 1. Hence the means test supports the conclusion that the subjects exhibited reciprocity in the investment game. The one-tailed Kolmogorov-Smirnov two-sample test does not detect that the treatment A amounts returned are significantly greater than the treatment B amounts returned (*p* = .135), although the Mann-Whitney test does detect a significant difference (*p* = .061).

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:

$$(1) \qquad R_t = \alpha + \beta D_t S_t + \gamma S_t + \varepsilon_t \,,$$

where

$$(2) \qquad D_t = 1 \text{ for treatment A data}$$
$$= 0 \text{ for treatment C data.}$$

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

(3) $R_t \in [0,3S_t]$.

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:

(4) $\sigma_t = \sigma e^{\theta S_t}$.

Note that $\hat{\beta}$ is the estimate of the effect of reciprocity on amounts returned by responders. We observe that $\hat{\beta}$ is positive and significantly greater than 0 ($p =$ .034); hence the tobit estimation supports the conclusion that the subjects exhibited positive reciprocity in the investment game.[5]

The means test, Mann-Whitney test, and tobit estimation all support the conclusion that there is positive reciprocity in this experiment, although the Kolmogorov-Smirnov test does not. I conclude, consistent with findings previously reported in the literature, that the zero-transfer prediction of the model of self-regarding preferences does fail empirically because subjects are motivated by positive reciprocity. But positive reciprocity is not the only reason for the traditional model's empirical failure because on average responders "return" 42% (= 100($2.06/$4.94)) as much money to proposers when the proposers have not undertaken any "friendly action" by sending money to the responders and, hence, there is no reason for responders to be motivated by reciprocity.

4. Concluding Remarks

This paper reports an experiment with a triadic design similar to that used in Cox (2000). The three games in the experiment are the investment game, a specific dictator game that controls for proposer motivations, and a specific dictator game that controls for responder motivations.

Several researchers had previously established 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 non-market fairness experiments, are inconsistent with the traditional model of self-regarding preferences. This leaves the profession with the task of constructing a 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 possible motivations. The experiments reported here make it possible to decompose transfers in the investment game into the parts motivated by altruism and the parts motivated by reciprocity or trust.

The triadic experimental design used here supports the following conclusions. There is significant trusting behavior in the investment game but, on average, 61% of the amounts sent by proposers to responders can be explained by proposers' altruistic other-regarding preferences. Similarly, there is significant reciprocity in the investment game but, on average, 42% of the amounts returned by responders to proposers can be explained by the responders' altruism. Therefore, if a theoretical model is to be capable of explaining behavior in the investment game it must include motivation by both altruistic other-regarding preferences over outcomes and internalized norms for responding to others' friendly actions (positive reciprocity) in its formal structure.

There are a few previous studies that have used some types of controls for discriminating between the observable implications of altruism and reciprocity. 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 but no significant difference between the rejection rates in the third party and standard games. Charness (1996) used Blount's control treatments in experiments with the gift exchange game. He found somewhat *higher* second mover contributions in the outside party and random treatments than in the standard gift exchange game. Bolton, Brandts, and Ockenfels (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 treatment and the positive and negative reciprocity treatments.

The general conclusion from the research reported in the present paper is about experimental methods: We cannot learn what we need to know in order to provide empirical guidance to theory development by designing single-game experiments to test theoretical predictions derived from the self-regarding preferences model. In order to provide informative data, we need to use designs that can discriminate between the observable implications of distinct motivations that can cause behavior to differ from that predicted by the traditional model.

Endnotes

* Financial support was provided by the Decision Risk and Management Science Program, National Science Foundation (grant number SES-9818561).

1. Some other papers in this research program are: Fehr and Falk (1999), Fehr, Gächter, and Georg Kirchsteiger (1996), and Fehr, Kirchsteiger, and Riedl, (1993).

2. This measure of transfers motivated by trust is formally derived in Cox (2000).

3. This measure of transfers motivated by reciprocity is formally derived in Cox (2000).

4. This procedure is followed in order to avoid any possible suggestion of "generalized reciprocity" (Yamagishi, forthcoming) to the responders, which would consist of repaying proposer C_j for the friendly action of proposer $A_{j.}$

5. Berg, Dickhaut, and McCabe (1995) noted that first movers who send all $10, and perhaps also those who send $5, might elicit larger returns by second movers. This possibility was examined for the data from the experiment reported here by introducing dummy variables for $5 and $10 amounts sent into the tobit model. Neither of the estimated coefficients for the $5 and $10 dummy variables was significantly different from zero.

References

Berg, Joyce, John Dickhaut, and Kevin McCabe, “Trust, Reciprocity, and Social History.” *Games and Economic Behavior*, July 1995, 10(1), pp. 122-42.

Blount, Sally, “When Social Outcomes Aren’t Fair: The Effect of Causal Attributions on Preferences.” *Organizational Behavior and Human Decision Processes*, August 1995, 63(2), pp. 131-44.

Bolton, Gary E., Jordi Brandts, and Axel Ockenfels, “Measuring Motivations for the Reciprocal Responses Observed in Simple Dilemma Games.” *Experimental Economics*, 1998, 1(3), pp. 207-19.

Cason, Timothy and Daniel Friedman, “Buyer Search and Price Dispersion: A Laboratory Study,” Working paper, University of California at Santa Cruz, 2000.

Charness, Gary, “Attribution and Reciprocity in a Simulated Labor Market: An Experimental Study.” Working Paper, University of California at Berkeley, 1996.

Cox, James C., “Trust and Reciprocity: Implications of Game Triads and Social Contexts,” Discussion Paper, University of Arizona, September 1999, revised July 2000.

Cox, James C., R. Mark Isaac, Paula Ann Cech, and David Conn, “Moral Hazard and Adverse Selection in Procurement Contracting,” *Games and Economic Behavior*, 17, 1996, pp. 147-76.

Cox, James C. and Ronald L. Oaxaca, “Laboratory Experiments with a Finite Horizon Job Search Model,” *Journal of Risk and Uncertainty*, 2, 1989, pp. 301-29.

Cox, James C. and Ronald L. Oaxaca, “Is Bidding Behavior Consistent with Bidding Theory for Private Value Auctions?”, in R. Mark Isaac (ed.), *Research in Experimental Economics*, vol. 6, Greenwich: JAI Press, 1996.

Cox, James C. and Ronald L. Oaxaca, “Good News and Bad News: Search from Unknown Wage Offer Distributions” (with Ronald L. Oaxaca), *Experimental Economics*, 2, No. 3, 2000, pp. 197-225.

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

Fehr, Ernst and Armin Falk, “Wage Rigidity in a Competitive Incomplete Contract Market.” *Journal of Political Economy*, 107, 1999, pp. 106-134.

Fehr, Ernst and Simon Gächter, “Fairness and Retaliation: The Economics of Reciprocity.” *Journal of Economic Perspectives*, Summer 2000b, 14(3), pp. 159-81.

Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger, “Reciprocal Fairness and Noncompensating Wage Differentials.” *Journal of Institutional and Theoretical Economics*, 152, 1996, pp. 608-640.

Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger, "Reciprocity as a Contract Enforcement Device: Experimental Evidence." *Econometrica*, July 1997, 65(4), pp. 833-60.

Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl, "Does Fairness Prevent Market Clearing? An Experimental Investigation." *Quarterly Journal of Economics*, May 1993, 108(2), pp. 437-60.

Güth, Werner, Nadège Marchand, and Jean-Louis Rulliere, "On the Reliability of reciprocal Fairness-An Experimental Study," Discussion Paper, Humboldt University of Berlin, 1997.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze, "An Experimental Analysis of Ultimatum Bargaining, *Journal of Economic Behavior and Organization*, 3, 1982, pp. 367-88.

Harrison, Glenn and Peter Morgan, "Search Intensity in Experiments," *Economic Journal*, 100, 1990, pp. 478-86.

Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir, "Bargaining and Market Behavior in Jerusalem, Ljublana, Pittsburgh, and Tokyo," *American Economic Review*, 81, 1991, pp. 164-212.

Smith, Vernon L., "Microeconomic Systems as an Experimental Science," *American Economic Review*, 72, 1982, pp. 923-55.

Smith, Vernon L. and Arlington W. Williams, "The Boundaries of Competitive Price Theory: Convergence Expectations and Transaction Costs," in L. Green and J.H. Kagel, eds., *Advances in Behavioral Economics*, vol. 2, Norwood, NJ: Ablex Publishing Corp., 1990.

Yamagishi, T., "Cross-Societal Experimentation on Trust: Comparison of the United States and Japan. In L. Ostrom (ed.), *Trust*, forthcoming.

Table 1. Decomposition Tests for Trust and Reciprocity

PARAMETRIC AND NONPARAMETRIC TESTS OF PROPOSER AND RESPONDER DATA					
Data	Send Mean	Return Mean	Means Tests	Kolmogorov-Smirnov 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 (.010)[1]	1.25 (.045)[1]	-2.35 (.010)[1]
Tr.A Return vs. Tr.C Return		…	2.88 (.018)[1]	1.00 (.135)[1]	-1.55 (.061)[1]
TOBIT ANALYSIS OF RESPONDER DATA					
$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$	LR Test	
4.20 (.060)	.680 (.034)[1]	-.759 (.124)	.158 (.008)	5.98 (<.025)	

p-values in parentheses.
[1] denotes a one-tailed test.
Standard deviations in brackets.
Number of subjects in braces.