VERIFIED TRUST:

RECIPROCITY, ALTRUISM, AND NOISE IN TRUST GAMES

MARIUS BRÜLHART
University of Lausanne & CEPR

JEAN-CLAUDE USUNIER
University of Lausanne

This version: October 2004

Abstract

Behavioral economists have come to recognize that reciprocity, the interaction of trust and trustworthiness, is a distinct and economically relevant component of individual preferences alongside selfishness and altruism. This recognition is principally due to observed decisions in experimental "trust games". However, recent research has cast doubt on the explanatory power of trust as a determinant of those decisions, suggesting that altruism may explain much of what "looks like" trust. Moreover, empirical tests for alternative behavioral determinants can be sensitive to experimental bias due to differences in protocols and framing. Therefore, we propose discriminatory tests for altruism and trust that can be based on within-treatment and within-subject comparisons, and we control for group attributes of experimental subjects. Our results support trust (i.e. expected reciprocation) as the dominant motivation for "trust like" decisions.

JEL classification: C91, D63, D64

Keywords: reciprocity, altruism, trust game, experimental error

Correspondence address: Ecole des HEC, University of Lausanne, CH – 1015 Lausanne, Switzerland. (Marius.Brulhart@unil.ch, Jean-Claude.Usunier@unil.ch)

Acknowledgements:

We are grateful for helpful comments received from Olivier Cadot and Mario Jametti as well as from seminar participants at the Universities of Lausanne and Nottingham. Hansueli Bacher, Tea Danelutti, Gregory Cleusix and David Viña have provided excellent research assistance. We thank the Institute of International Management of the University of Lausanne (IUMI), and the Swiss National Science Foundation for financial support.

VERIFIED TRUST: RECIPROCITY, ALTRUISM, AND NOISE IN TRUST GAMES

Trust, but verify.

Russian proverb

1. INTRODUCTION

Mainstream economic theory is built squarely on the model of individuals as rational maximizers of own utility, with a rare subsidiary role conceded to altruistic concerns for "fairness" or equality. In an exciting recent development, this basic theoretical building block has been refined to take account of observed behavioral regularities that depart from the pure *homo oeconomicus* paradigm. Even where theory had formerly allowed for non-selfish motivations, utility had been defined strictly over *outcomes* of individual actions. A strong argument has been made, however, that *intentions* matter too: kind deeds are reciprocated with kind deeds, and unkind deeds with unkind ones, even in pure one-shot situations. Such behavior cannot be traced back to pure individual self interest. Thus, *reciprocity* has come to be seen as a third major determinant of economic behavior, in addition to selfishness and altruism.

Positive reciprocity, i.e. "reciprocal kindness", is equivalent to the combination of *trust* and *trustworthiness*. We focus on the measurement of trust.[1] Because of its value for relationships where formal contracting is costly, trust has been called a "lubricant" for the market economy (Arrow, 1974a), it has been shown to affect optimal contract and institutional design in a range of economic situations (Fehr and Gächter, 2000; Fehr and Fischbacher, 2002), and it has been found to favor the formation of large firms and organizations (La Porta, Lopez-de-Silanes, Shleifer and Vishny, 1997), to promote

[1] A number of researchers have explored the determinants of trustworthiness (Ben-Ner, Putterman, Kong and Magan, 2004; Clark and Sefton, 2000; Charness and Haruvy, 2002; Cox, 2003; Cox and Friedman, 2002; McCabe, Rigdon and Smith, 2003; Nelson, 2002). The main difference between laboratory analyses of trust and of trustworthiness is that the latter can draw on *observed* actions by trustors, while the former must incorporate *expectations* about trustees' trustworthiness.

international investment and trade (Guiso, Sapienza and Zingales, 2004) and to increase economic growth (Zack and Knack, 1999).

This paper is about ways to distinguish trust empirically from other motivations. Much empirical analysis of trust has been based on survey answers to questions such as "generally speaking, would you say that most people can be trusted?". This type of evidence is valuable for cross-cultural research (see e.g. Alesina and La Ferrara, 2002), but there are inevitable limits to the accuracy and comparability of survey-based information (see, e.g., Glaeser, Laibson, Scheinkman and Soutter, 2000). More rigorous empirical evidence in favor of trust and reciprocity has come from laboratory experiments. The "trust game" of Berg, Dickhaut and McCabe (1995) has become a standard experiment to measure trust. In the trust game, a *first mover* is randomly and anonymously paired with a *second mover*, both are given a monetary endowment, the first mover may transfer some or all of his endowment to the second mover, this transfer is tripled by the experimenter and handed to the second mover, and finally the second mover may return some or all of the first mover's transfer.[2] First-mover transfers are interpreted as a manifestation of trust, and second-mover transfers as a manifestation of trustworthiness.[3]

The interpretation of transfers in that game as manifestations uniquely of trust and trustworthiness has recently been questioned, because *altruism* might also play a role in non-zero ("trust-like") transfers. A number of researchers have compared transfers in trust games to transfers in comparable dictator games

[2] The seminal trust-and-reciprocity experiment is the "gift-exchange game" of Fehr, Kirchsteiger and Riedl (1993). The main differences between the gift-exchange game and the trust game are, first, that the first-mover transfers in the gift-exchange game are determined through a bidding process (where above market-clearing transfers signal "kindness"), and, second, that, in the gift-exchange game, the positive-sum element appears at the level of *second*-mover transfers (which are multiplied by the experimenter using a convex schedule). An interesting extension of the trust game has been put forward by Abbink, Irlenbusch and Renner (2000). Their "moonlighting game" has the same structure as the trust game but allows for negative as well as positive reciprocity by the two players.

[3] Conflicting results exist on the correspondence between responses to the typical attitudinal survey question on trust, and transfers made in trust games. In a sample of Harvard undergraduate students, Glaeser *et al.* (2000) observe no correlation between first-movers' survey answers and their transfers made. Conversely, Fehr, Fischbacher, von Rosenbladt, Schupp and Wagner (2003), in a design that mixes survey and experimental methods

and found the differences to be surprisingly small. Cox (2003), for example, finds that dictators sent 70 percent of the amounts transferred by equally endowed trust-game players.

We argue that the interpretation of trust-game transfers is complicated by a further factor, "noise". This factor captures idiosyncrasies in individual preferences and, more importantly, potential biases induced by the framing and practical implementation of experiments. The mere fact that a game is based on a single choice (such as the dictator game) or on a sequence of choices (such as the trust game) may influence transfers, and difficult-to-control details of practical implementation can introduce treatment-specific biases. This is particularly relevant in games with low stakes, where it may be rational for agents not to think through the full material payoff structure.

Treatment-specific biases in the noise term are innocuous as long as comparisons are restricted to observations drawn from the same experimental treatment. Hence, we develop a discriminatory criterion that allows to test for the significance of altruism as a determinant of "trust-like" first-mover transfers *within treatments*. We do this by differentiating second movers by their experimental endowment and test whether first movers transfer more to "poor" second movers than to "rich" ones. In one of our experimental treatments, we furthermore adopt a *within-subject* design by letting each first mover play simultaneously with a poor and a rich second mover, so that we can control for potential individual-specific biases as well as for treatment-specific bias. Using data gathered in three experimental sessions with undergraduate university students, we do not find a significant negative relationship between first-mover transfers and second-mover "wealth" in any of our regression specifications or experimental treatments. Our results therefore reject the hypothesis of altruistic motives as statistically significant determinants of "trust-like" decisions.

Another feature of our study is to use post-experimental questionnaires to refine the analysis of observed choices. Answers to a question on first movers' expected returns allow us to test the original

---

for a broad cross-section of German residents, find that survey responses have strong predictive power for first-mover trust-game transfers.

interpretation of observed trust-game decisions, that is whether "trust-like" transfers are indeed significantly affected by expected reciprocation. Even though questionnaire-based data on expectations are prone to measurement error, and estimated coefficients on those measured expectations are thus biased downwards, we find significantly positive coefficients when regressing first-mover transfers on expected second-mover returns.[4] These results confirm the hypothesis that "trust-like" behavior is indeed motivated by trust, i.e. the expectation of reciprocal kindness on the part of second movers.

Finally, information gleaned from the questionnaires allows to test for group specific differences in trust-game transfers. We find that economics students are significantly less trusting than non-economics majors, but that gender, nationality, mother tongue and experimental treatment have no statistically significant impact on first-mover transfers.

The paper is structured as follows. Section 2 reviews the literature on laboratory-based measurement of trust. A behavioral model of first-mover motivations in trust games is developed in Section 3, and discriminating hypotheses on altruism and reciprocity are derived. The experimental protocol is described in Section 4. Section 5 reports our experimental findings and tests the two discriminating hypotheses econometrically. Section 6 concludes.

2. PLAYING GAMES WITH ALTRUISM AND RECIPROCITY: THE LITERATURE

2.1 The Starting Point: Behavior in Trust Games as a Manifestation of Reciprocity

By their very definition, selfish and reciprocal motivations are compatible with perfect anonymity of the interacting individuals (Jencks, 1990; Fehr and Gächter, 2000). One can behave perfectly selfishly *vis-à-vis* a total stranger, and one may reciprocate a friendly or unfriendly action even if nothing else is known

[4] Expected second-mover returns are expressed as shares of first-mover transfers in order to minimize potential simultaneity bias.

about the individual concerned. In contrast, altruism is widely considered to be "context dependent" (Eckel and Grossman, 1996), i.e. it is negatively related to social distance (Jencks, 1990; Bohnet and Frey, 1999).

In the double-blind one-shot trust game experiments, transfers made by first and second movers are incompatible with a society in which all agents behave purely selfishly and all agents expect all other agents to behave purely selfishly. Strict anonymity of experimental subjects is imposed in order to rule out altruistic motives or reputation effects, and to leave only reciprocity as a motivational force in addition to selfishness. In this view, first-mover transfers measure trust and second-mover transfers measure reciprocation. Agents in such games have consistently made significantly positive transfers in both directions. This has been interpreted as strong evidence for the pervasiveness of reciprocity as a motivator of social behavior, and it allows for interesting intercultural comparisons.[5] In the United States, Berg *et al*. (1995) observed that first movers on average entrusted 51.6% of their endowment to second movers. Replicating that game in France and Germany, Willinger, Keser, Lohmann and Usunier (2003) found French students to be less trusting than Germans, the former transferring on average 42% of their endowment, compared to the 66% entrusted by the representative German subject. Fershtman and Gneezy (2001) concluded that Israeli subjects considered Ashkenazic second movers more trustworthy than Sephardic second movers, since average first-mover transfers corresponded to 76% and 40% of first-movers' endowments respectively. Bornhorst, Ichino, Schlag and Winter (2004), playing the game with PhD students, observed that northern Europeans trust more, and thus are trusted more, than southern Europeans. Finally, Fehr and List (2004) found that managers are more trusting than university students: in their experiments (conducted in Costa Rica) CEOs sent 59% on average of their initial endowment, while students sent 40%.

[5] For a survey, see Camerer (2003, ch. 2.7).

2.2 The Challenge: Behavior in Trust Games as a Manifestation of Altruism

In order to interpret non-zero transfers in standard trust games as measures of reciprocity, one has to assume that altruistic motives cannot exist in an anonymous setting. One could, however, argue that the standard design does not in fact grant perfect anonymity. Subjects, who are traditionally undergraduate university students, know that their counterparts are drawn from the same population. Students might perceive their social distance to other students, even if randomly chosen, to be small enough for them to qualify for altruistic treatment (as if they were all members of the "family of University X students"). Alternatively, they might feel what Jencks (1990) has termed "moralistic unselfishness". This is a form of altruism that extends even to individuals with whom one has no direct contact, no prospect of direct contact and no particular emotional connection through some shared features (such as ethnicity, gender, age, etc.) except for the empathic feeling of being members of the same species. Such altruism is compatible with sharing some of the spoils of an experiment with an unknown fellow student.[6]

The possibility that transfers in trust games may be motivated by a mixture of altruism and reciprocity has been recognized by Smith (2003). He reports on a three-node extended trust game, in which, at the initial decision node, the first mover has a choice between a conclusive payoff that is very favorable to the second mover (call it the "altruistic allocation") and a continuation of the game to a node from where the ultimate outcome will increase in the degree of reciprocity between the two players but the second mover cannot reach a payoff as high as under the altruistic allocation. Strong altruism would advocate that the game ends at the initial node, with first movers choosing the allocation most favorable to second movers. None of Smith's 26 experimental subjects makes this choice. This evidence rejects altruism as a *dominant* determinant of transfers in trust games. However, it cannot rule out altruism as *part of* the motivation underlying "trust-like" giving (altruism just is not strong enough for first movers to make the

[6] One might object that if students were prepared to make positive transfers in anonymous trust games because of altruistic motives, they should also be prepared to leave banknotes randomly on the campus. However, it is plausible to think that the opportunity cost in utility terms of sharing part of an experimental windfall when agents are explicitly offered that option is substantially lower than that of scattering earned money around the campus

sacrifice implied in the altruistic allocation). Moreover, opting for the altruistic allocation in this game is in fact incompatible with altruism defined as inequality aversion (see e.g. Fehr and Schmitt, 1999), since first-movers' alternative option at the initial decision node gives them access to more equal allocations further down the decision tree.

Another approach to test for altruism has been developed by Cox (2001, 2003, 2004). He proposes a "triadic" experiment where one group plays the standard trust game and two similarly sized control groups play dictator games. Members of the first group of dictators are given amounts that are equal to those allocated to first movers in the investment game. Hence, all dictators in that group are endowed with the same amounts. Dictators in the second control group are given amounts that are equal to those at the disposal of second movers in the trust game inclusive of the transfers received from first movers. Hence, dictators in the second group are not all endowed with the same amounts.

Cox interprets transfers made by dictators as being motivated by altruism, which leads him to attribute the difference between transfers observed in the trust game and transfers observed by the respective control-group dictators as a measure of reciprocity. Running the experiment several times with University of Arizona students, he finds that control-group transfers amount to between 61 and 97 percent of first-mover trust-game transfers. This would suggest that a major share of what has commonly been interpreted as trust-based transfers is in fact motivated by altruism.

Similar evidence is found in several recent studies. Buchan, Croson and Dawes (2001) carry out (*a*) standard trust games, and (*b*) amended trust games in which second-mover transfers are not given to the first movers from which the second movers have received their transfers, but to a randomly chosen first mover. The amended trust game is effectively a two-way dictator game. They find that first-mover transfers in the amended game amount to 65 percent of transfers in the standard trust game. Dufwenberg and Gneezy (2000) compare dictator transfers to second-mover transfers made in an experiment that

(where the evident alternative is to give money to some identified recipient). Andreoni and Miller's (2002) finding that altruistic choices are price sensitive can be enlisted in support of this conjecture.

resembles the trust game, and they find no significant difference. McCabe, Rigdon and Smith (2003) report results of an experiment that closely resembles the second-mover part of Cox's triadic setup, the main difference being that they allow only two possible transfers - five or nothing. They observe that 33% of second movers transfer five in the dictator-game treatment and 65% of second movers transfer five in the trust-game treatment. Fershtman and Gneezy (2001) compare first-mover trust game transfers to dictator transfers in an experimental design that breaks the anonymity of subjects vis-à-vis the experimenter (because their aim is to study discrimination in Israeli society). From the mean transfers they report, one can calculate that dictator transfers correspond to up to 69% of first-mover transfers in the trust game. Finally, Charness (2004) reports on a gift-exchange game using the protocol of Fehr *et al.* (1993) but adding a control treatment where first-mover transfers ("wages") are randomly created rather than offered by the first movers ("employers"). Charness (2004) finds that mean second-mover transfers ("effort") are actually *higher* in the control treatment than in the standard gift-exchange treatment.

The high average transfers by control-group dictators compared to trust-game subjects have led some experts to question the strength of the trust-reciprocity hypothesis. Surveying this literature, Camerer (2003, p. 100), for example, concludes that "repayments are mostly the result of altruism and are increased only a little by reciprocation".

2.3 A Further Complication: Noise

In addition to trust and altruism, a realistic theory of experimental behavior must include randomness as a possible explanation for non-zero transfers by trust-game subjects. Even if subjects' decisions in the laboratory happened to coincide *on average* with their hypothetical preferred choice under perfect information, sufficient stakes and neutral framing, observed decisions will have a random component and thus not reflect the hypothetical preferred choice in each case. Andreoni (1995b, p. 893) points out that "subjects [may] have somehow not grasped the true incentives", and calls this effect "confusion". Smith (2003, p. 494) considers the possibility that "subjects are game-theoretically unsophisticated".

Not to grasp the incentives of the game fully could of course be a rational decision by experimental subjects who weigh up the intellectual effort of carefully considering their options against the potential returns. This view of costly information processing by experimental subjects is corroborated by the result, found across a number of different games, that raising the stakes, while mostly neutral on mean transfers, significantly reduces the variance of observed decisions (for a survey, see Camerer and Hogarth, 1999).

In addition, it is well documented that experimental transfers are sensitive to the wording of instructions (Andreoni, 1995a; Bolton, Katok and Zwick, 1998; Burnham, McCabe and Smith, 2000; Charness, Frechette and Kagel, 2004; Hoffman, McCabe and Smith, 1996). Subtle differences in the language of instructions sheets can significantly change the amounts transferred. Their mere inexperience with the experimental situation can make subjects sensitive to small procedural features and thus bias the results, depending on which way the framing effects depart from perfect neutrality. Moreover, it may be that the very fact that subjects are given the option to make a transfer predisposes them towards thinking how much they should transfer, and not whether they should transfer at all. In that case, randomness takes the form of positive "experimental bias".

The implication of randomness is that individual transfers observed in trust and dictator games must be interpreted as noisy measures of trust and altruism. In particular, if experimental bias exists, it is no longer possible to determine the relative magnitude of reciprocity and altruism as motivational forces based on the differences between transfer levels of trust-game subjects and their peers in the dictator-game control group.

Cox (2001, 2003, 2004) and Buchan *et al.* (2001) take account of randomness by computing statistical significance tests on the difference in mean transfers between trust-game treatments and control treatments, and they find that trust-game transfers are statistically significantly higher. This supports the trust-reciprocity hypothesis. Since their experiments for different treatments were conducted sequentially, one might however argue that framing effects could have differed (the protocols differ, and

it just takes one wrong word by an experimenter) and/or that information could have flowed outside the laboratory among participants of different sessions.[7] Note that different behavioral norms could be triggered by varying experimental protocols. For example, it may well be that subjects make some transfers for the mere reason that the option of making a transfer is given to them, or out of sheer curiosity about what will happen to them when the game has more than one decision node.[8] For valid statistical inference, one would therefore wish to make within-session rather than between-session comparisons, with experimental protocols held as similar as possible, so that any experimental biases would affect treatment and control groups identically and thus wash out in the comparison.

3. ALTRUISM VERSUS TRUST VERSUS NOISE: DERIVATION OF DISCRIMINATING HYPOTHESES

3.1 First-Mover Transfers in Trust Games: A Behavioral Model

In a quest for analytical rigor and transparency of underlying assumptions, we propose a model of subject motivations in trust games. The model is kept as parsimonious as possible while incorporating the key behavioral elements put forward in the literature.

The trust game can be formally described as follows. First movers i start the game with a money holding of y_i. Second movers j have an initial money holding of y_j. At the first stage of the game, first movers can send any amount s_i, $0 \leq s_i \leq y_i$, to their paired second movers.[9] The experimenter triples the amount sent, so that second movers receive $3s_i$. At the second stage of the game, second movers can return any

[7] Cox (2004, p. 270) gives a description of the framing issue, raised by a referee: "The argument was that, while the games in the three treatments may look similar using the author's [i.e. Cox'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."

[8] This is particularly relevant for the "triadic" game, since curiosity about later stages may play a role in the trust game but not in the dictator game.

[9] We abstract here from the fact that amounts sent in experiments must take discrete values.

amount r_j, $0 \le r_j \le (y_j + 3s_i)$, to their paired first movers. We call "rate of return" the ratio $\rho_j = r_j / s_i$; and we denote holdings at the end of the game by Y_i and Y_j, for first and second movers respectively. Hence, $Y_i = y_i - s_i + r_j = y_i + s_i(\rho_j - 1)$, and $Y_j = y_j + 3s_i - r_j = y_j + s_i(3 - \rho_j)$. Finally, players' beliefs about actions of others are denoted with a circumflex. Thus, we write first movers' expected rate of return from second movers as $\hat{\rho}_j$.

A General Utility Function

We specify the following general expected utility function of an agent in a two-player sequential game that is restricted to strictly non-negative transfers:

$$U_i = E(\, f(Y_i, Y_j, [K_i, K_j], A_i\varepsilon_i)), \qquad \textbf{(1a)}$$

where f is continuous and twice differentiable in its four arguments, K represents "kindness" (to be defined below), A stands for the agent's own action, and ε is a mean-zero stochastic term. Furthermore,

$$\frac{\partial U_i}{\partial Y_i} \ge 0, \qquad \frac{\partial^2 U_i}{\partial Y_i^2} \le 0, \qquad \frac{\partial\left(\frac{-U_i''}{U_i'}\right)}{\partial Y_i} \le 0, \qquad \textbf{(1b)}$$

$$\left.\frac{\partial U_i}{\partial Y_j}\right|_{Y_j<Y_i} \ge 0, \quad \left.\frac{\partial U_i}{\partial Y_j}\right|_{Y_j<Y_i} \ge \left.\frac{\partial U_i}{\partial Y_j}\right|_{Y_j>Y_i}, \quad \frac{\partial^2 U_i}{\partial Y_j^2} \le 0, \qquad \textbf{(1c)}$$

$$\frac{\partial U_i}{\partial [K_i, K_j]} \ge 0, \text{ and} \qquad \textbf{(1d)}$$

$$\frac{\partial U_i}{\partial [A_i\varepsilon_i]} \ge 0. \qquad \textbf{(1e)}$$

Assumptions (1b) represent standard concave preferences over own income with nonincreasing absolute risk aversion. Assumptions (1c) define altruism, incorporating an element of inequality aversion: my

utility gain from a given increase in your payoff is larger if this makes you less poor than I than if this makes you even richer than I.[10] Assumption (1d) represents "intrinsic reciprocity" (Sobel, 2004): irrespective of final outcomes, agents' utility increases if they feel treated kindly and if they can reciprocate kindly (vice-versa for unkind actions). Note that trusting behavior could either be motivated entirely by intrinsic reciprocity (a "desire to elicit kindness through kindness"), or by a combination of own-payoff maximization (selfishness) and expected intrinsic reciprocity on the part of the other agent. Finally, (1e) expresses that agents' utility may be affected by idiosyncratic factors, which we attribute to experimental error.

In order to apply the general utility function (1a) to an analysis of first-mover transfers in a laboratory trust game, two issues require us to impose further structure.

First, the first two arguments of utility function (1a-e) are defined over end-node outcomes. Hence, we need to model first-movers' beliefs about second-movers' reactions. We make two simplifying assumptions:

- **(1f)**: First movers reason about second-movers' returns in terms of *rates* of return. Specifically, they hold beliefs about the expected rate of return ($\hat{\rho}_j$) and the variance thereof, both of which are independent of the amount sent (s_i).
- **(1g)**: In so far as their altruistic motive is concerned, first movers abstract from possible second-mover strategies. Hence, altruism, if present, is guided solely by payoffs at the first decision node, $y_j + 3s_i$ and $y_i - s_i$.[11]

These assumptions, while intuitively rather plausible, are not innocuous. They imply bounded rationality and allow us to abstract from higher-order conjectures by first movers.

[10] Our specification encompasses the possibility that, if you are richer than I, an increase in your payoff reduces my utility.

[11] This assumption could represent bounded rationality on the part of first movers who look at initial relative payoffs as a sufficient approximation of end-node relative payoffs.

Second, in order to quantify kindness, we need to define a benchmark action that represents zero kindness.

- **(1h)**: For first movers, the no-kindness benchmark is at $s_i = 0$, the subgame perfect strategy of purely selfish agents. For second movers, the benchmark is $\rho_j = 1$, above which any return implies kindness.

Hence, kindness implies a positive first-mover transfer (as it exposes first movers to the risk of losing out at the expense of second movers) and a second-mover transfer that returns more than what first movers sent (as it rewards first movers with a positive return for their risky strategy that has benefited second movers).

Finally, in view of later empirical application, we narrow down the preference model further, by imposing the following two assumptions:

- **(1i)**: f is additive in its arguments.
- **(1j)**: ε_i is a random variable drawn independently from the same distribution for every agent i. The distribution of ε_i is normal around a mean X ($\varepsilon_i \sim N(X, \sigma_\varepsilon^2)$), and it is uncorrelated with y_j and $\hat{\rho}_j$.

We allow for a potentially non-zero mean of the idiosyncratic term in (1j) because confusion and framing effects may bias s_i upward.[12]

Utility Function of Laboratory "Trustors"

Based on (1a-j), we can write the following utility function of first movers at the first decision node of the standard laboratory trust game:

[12] Strictly speaking, our model therefore implies that $E(\varepsilon_i) > 0$; but for the analysis that follows it suffices to impose the weaker restriction $E(\varepsilon_i) \neq 0$.

$$U_i = a*E(f^{selfish}(\, y_i + s_i[\hat{\rho}_j - 1]\,)) + b*(f^{altruist}(\, y_j + 3\, s_i,\, y_i - s_i)) + c*E(\, s_i \hat{\rho}_j - s_i^{\lambda}\,) + \varepsilon_i s_i, \qquad \textbf{(2)}$$

$$\lambda > \hat{\rho}_j\,.$$

Assumptions (1b-e) carry over to the four arguments of (2). Hence, (1b) and (1c) define the admissible functional forms for $f^{selfish}$ and $f^{altruist}$. The specific functional form that we impose on the third argument of (2) ensures that first movers, to the extent that they are motivated by intrinsic reciprocity, will transfer more (i.e. be kinder) the more they expect second movers to return (i.e. the more kindly they expect second movers to react), conforming with assumptions (1d) and (1h). The first and third arguments of (2) are written in terms of expected utility, since $\hat{\rho}_j$ is a random variable from the point of view of first movers.

3.2 Two Discriminating Hypotheses

We have identified four determinants of first-mover transfers in the trust game, but we only observe one variable s_i. The challenge is to design the experiment such that separate determinants of s_i become identifiable.

Our first approach is to vary second movers' initial wealth y_j, and to examine the relationship between s_i and y_j. (2), (1c) and (1f) imply that:

$$\frac{\partial s_i}{\partial y_j} < 0 \quad \text{iff} \quad b > 0, \quad \text{and} \qquad \textbf{(3a)}$$

$$\frac{\partial s_i}{\partial y_j} = 0 \quad \text{iff} \quad b = 0. \qquad \textbf{(3b)}$$

If altruism features, first movers will give more to poor second movers than to rich ones. Conversely, according to (3b), first-mover transfers in the absence of own altruistic motives will be unrelated to

second-mover wealth.[13] (3b) represents the behavioral model implicit in the original interpretation of trust-game results. Expressions (3) thus provide a crisp discriminating hypothesis.

Proposition 1:

Altruism implies that first movers send more to poor second movers than to rich second movers. In the absence of altruism, first-mover transfers are unrelated to second-mover wealth.

Proposition 1 has the advantage of being based entirely on directly observable variables. Its drawback is that it is not an explicit test of the relevance of *trust* as a motivating force underlying first-mover transfers. If we are prepared to extend the analysis to expected reciprocation $\hat{\rho}_j$, a variable that cannot be observed directly, we can postulate the following:

$$\frac{\partial s_i}{\partial \hat{\rho}_j} = 0, \qquad \text{iff} \quad a = 0, \quad c = 0, \text{ and} \tag{4a}$$

$$\frac{\partial s_i}{\partial \hat{\rho}_j} > 0 \qquad \text{otherwise.} \tag{4b}$$

Hence, we can formulate a second discriminatory proposition.

Proposition 2:

Trust implies that first-mover transfers increase in expected second-mover returns. In the absence of trust, first-mover transfers are unrelated to expected second-mover returns.

[13] Note that we have not determined how first-movers' expectation $\hat{\rho}_j$ is formed, hence implying that it is based on a universally shared heuristic norm. If we allowed for more fully rational expectation formation, $\hat{\rho}_j$ might plausibly correlate positively with y_j, as richer second movers with concave utility over own payoff would be "better able to afford" reciprocation, and their altruistic motive too would favor more generous reciprocation. In that case, the first derivative in (3b) would be positive. Furthermore, the qualitative result that s_i increases in y_j

As (4a) shows, trust has two components: intrinsic reciprocity and selfish own-payoff maximization. With intrinsic reciprocity taking the form stipulated in (2), it is easy to derive that s_i increases in $\hat{\rho}_j$: the kinder I expect you to be, the more kindly I feel like treating you. If first movers hold no intrinsically reciprocal preferences but expect second movers to hold them, their selfish motive will still motivate first-movers to make transfers. This situation in fact corresponds to the textbook asset allocation problem of a risk-averse agent choosing between a safe asset (keep the money) and a risky asset (make a transfer to the second player). As shown by Arrow (1974b, p. 105), concave utility over own payoff with decreasing absolute risk aversion, as assumed in (1b), implies that first movers' demand for the risky asset (s_i) will be positively correlated with additive shifts in the expected return ($\hat{\rho}_j$).

3.3 Testing for Altruism

We can now specify the following baseline model for estimation of Proposition 1:

$$s_i = C + \beta y_j + e_{0i}, \qquad e_{0i} = \varepsilon_i - E(\varepsilon_i), \qquad e_{0i} \sim N(0, \sigma_\varepsilon^2), \qquad C = E(\varepsilon_i) + \Omega, \qquad \textbf{(5)}$$

where C, β and e_{0i} are unobserved. Ω stands for transfers motivated by trust. If we use OLS, assumption (1j), combined with the normality of the distribution of e_{0i}, implies that we will obtain unbiased estimates of C, β and e_{0i}, and that standard inference can be applied. Note that C is a biased estimate of Ω, which is why the comparison of mean transfers in the triadic experiment is problematic.

Our discriminating criterion can be evaluated through a t test on the null hypothesis that $\beta_{OLS} = 0$. If the null hypothesis is rejected, and β_{OLS} is negative, then we infer that altruism plays a significant role in determining s_i. If the null hypothesis cannot be rejected, then we conclude that, relative to the noise in

---

would hold even if altruism were linear, i.e. if the last derivative in (3a) were equal to zero. In that case, s_i would be bigger if $y_j < y_i$ than if $y_j > y_i$ but constant for variations of y_j in each of those ranges.

our data, altruism is not a significant determinant of s_i, and s_i thus represents a combination of trust and experimental error. Finally, if the null hypothesis is rejected and β_{OLS} is positive, our model is rejected by the data.[14]

We consider three extensions to this baseline empirical model. First, we allow some variation in agents' trust motives. Specifically, we now maintain that Ω and/or $E(\varepsilon_i)$ can differ across population groups. These groups can be characterized by criteria such as gender, nationality, educational background or date of the experiment. This extension allows for the possibility that, despite our randomized experimental design, y_j happens to be correlated with some grouping that affects s_i, in which case the baseline estimate of β would be biased.[15]

Suppose, for example, we consider only a single criterion, gender, represented by a dummy variable G_i, set to 1 for women. Our empirical model thus becomes:

$$s_i = C + \delta G_i + \beta y_j + e_{1i}. \qquad \textbf{(6)}$$

Adding additional grouping criteria would simply add further intercept-shifting parameters δ to the model.

Second, we also allow some variation in agents' altruistic motives, since the altruism motive could be significant for some groups and not for others. Using again the example of grouping by gender, the model becomes:

$$s_i = C + \delta G_i + \beta_0 y_j + \beta_1 G_i y_j + e_{2i}, \qquad \textbf{(7)}$$

[14] An extension of the model along the lines sketched in footnote 10 could accommodate a positive coefficient.

[15] This is possible only in the between-subjects version of our experiment.

where β_1 captures the differential effect on altruism-induced transfers of female compared to male players.[16]

Third, we take account of the fact that the trust game imposes bounds on the dependent variable s_i. To correct for potential bias arising through this double censoring, we use a two-limit Tobit estimator, with censuring points corresponding to the experiment-specific lower and upper bounds on s_i.

3.4 Testing for Reciprocity

Our empirical test developed above pits a null hypothesis of *altruism* against the alternative of no altruism. Based on Proposition 2, we can formulate a complementary test, still concerning s_i but setting up a null hypothesis of *trust-reciprocity* versus an alternative of no such motivation. The reciprocity-augmented version of the group-wise model (7) becomes:

$$s_i = C + \phi \hat{\rho}_j + \delta G_i + \beta_0 y_j + \beta_1 G_i y_j + e_{3i}, \qquad \textbf{(8)}$$

where C, δ and β are the same as in (7). A test for reciprocity simply means comparing the null hypothesis $\phi = 0$ with the alternative $\phi > 0$. Rejection of the null hypothesis in favor of the alternative implies significance of the trust-reciprocity motive. If both the null and the alternative hypotheses are rejected, i.e. $\phi_{OLS} < 0$, our model is rejected by the data.

Why do we not incorporate $\hat{\rho}_j$ in our regression specification from the start? The reason is that $\hat{\rho}_j$ is neither a design feature of the experiment (like y_j) nor an observable strategy chosen by subjects (like s_i),

[16] When, as in our example, there is only one grouping variable, then separate regressions for the each group would be equivalent to estimating equation (8). When we control for multiple overlapping groupings, however, the interaction specification *à la* equation (8) is different from, and superior to, group-wise regressions.

as it can only be observed by asking subjects.[17] Model (8) therefore mixes experimental with survey methods. Given that the former has been developed as a way to reduce the informational imprecision typically associated with the latter, moving from (7) to (8) implies the concession of some observational accuracy. $\hat{\rho}_j$ being measured with error, the ϕ_{OLS} and its associated standard error will be unambiguously downward biased (see e.g. Meijer and Wansbeek, 2000). Since we have no perfect palliative for this problem, our test for reciprocity is biased in favor of acceptance of the null hypothesis of no reciprocity. The odds of our test are thus stacked against diagnosing trust.

4. EXPERIMENTAL PROTOCOL

We have played the trust game with undergraduate students at the University of Lausanne. First movers were all endowed with y_i = 10 Swiss francs.[18] Second movers were differentiated by the size of their show-up fee y_j, some starting the experiment with nothing, some with 10 francs and some with 20 francs. First movers knew the size of y_j of their paired second movers, and second movers knew their paired first movers' endowment y_i.

We played this game in three sessions, using standard double-blind procedures. No subject had participated in an experiment before, none played more than once, and they were allocated randomly to first- or second-mover roles. Subjects were recruited by email sent to all University of Lausanne first-year undergraduate students.[19]

[17] Note that $\hat{\rho}_j$ and y_j are uncorrelated in our behavioral model, which leaves β_{OLS} of regression equations (6) to (8) unbiased even though $\hat{\rho}_j$ is omitted from the estimations. In the empirical part, we explicitly test for omitted-variables bias (and can reject it in all cases).

[18] At the time of the experiments, one Swiss franc was worth approximately 0.73 US dollars.

[19] The texts of the "recruitment email", experimental instruction sheets and the post-experiment questionnaire can be obtained from the authors on request.

- **Session A** was played manually with physical money (coins of 1 franc) on January 20, 2003. 38 first movers played with one second mover each. 13 second movers started the game with nothing, 12 started with 10 francs and 13 started with 20 francs.
- **Session B** was played manually with physical money (coins of 1 franc) on January 22, 2003. 18 first movers played with two different second movers each, one with no show-up fee, and one with a show-up fee of 20 francs.
- **Session C** was played via internet during the first week of June 2003. 31 first movers played with one second mover each. 16 second movers started the game with nothing, and 15 started with 20 francs.

For sessions A and B, we used standard manual procedures, with first movers, second movers and experimenters in separate rooms and money circulating physically in sealed envelopes. Before leaving the venue of the experiments, subjects were asked to fill in a questionnaire which did not compromise their anonymity.[20]

Session C was conducted using a novel web- and email-based protocol managed by an independent monitor so as to respect anonymity among players and vis-à-vis the experimenter. Subjects were recruited via email and retained if they had not previously taken part in session A or B. First movers were randomly selected and invited by email to go to a web page with the relevant instructions. They were attributed an individual code allowing the monitor to match transfers. The second stage started two days later, once all first movers had made their decisions. Second movers were then invited to connect to a web page with their instructions. There, they also learned their initial endowment (0 or 20 francs) and the amount that had been sent by their paired first movers and tripled by the monitor. Second movers were asked to make their return decisions and to fill out the post-experiment questionnaire on the web page. Once second movers had made their decisions, an email was automatically sent to their paired first movers. First movers were than asked to fill out a post-experiment questionnaire. Finally, all players

[20] Ortmann, Fitzgerald and Boeing (2000), comparing treatments with and without questionnaires, report that the introduction of anonymity-preserving questionnaires in trust games has no significant impact on transfers made.

were invited to collect their earnings from an administrative clerk. Players had no way of finding out each other's identity, since they exchanged mails only with the monitor. The experimenters were also kept uninformed about the identity of the players.

The manual sessions required about two hours, whereas the computerized session took five days, due to the sequential and decentralized experimental setup. Descriptive statistics on the composition of the subject pools and on transfers made are given in Table 1.

We chose to conduct the experiment in three different settings as a robustness check of our inference with respect to the experimental protocol used. This allows us to compute between-treatment differences as well as within-treatment differences, i.e. we can compare transfers made by different subjects in exactly the same experimental setting and with identical instructions. We can thereby eliminate the potential bias that exists when observed transfers are compared across different sessions and/or protocols. In Session B, using a within-subject design, we control not just for possible treatment-specific experimental biases, but for subject-specific biases. We thus expect Session B results to be particularly statistically powerful.[21]

5. EXPERIMENTAL RESULTS

Summary statistics of observed transfers are reported in Table 1 and illustrated in Figure 1. We find that both first movers and second movers made large transfers in all three sessions. First movers on average sent 7.02 of their 10 francs to second movers, and second movers on average returned 10.32 francs. As a point of comparison, two trust games played with US students and both y_i and y_j of 10 dollars yielded

[21] Camerer (2003, p. 42) notes that "there is a curious bias against within-subjects designs in experimental economics", and that "one possible reason is that exposing subjects to multiple conditions heightens their sensitivity to the differences in conditions. This hypothesis can be tested (...) by comparing results from within- and between subjects designs, which is rarely done." Since heightening first movers' sensitivity to y_j is precisely

averages sent (returned) of 5.16 (4.66) dollars (Berg *et al.*, 1995) and 5.97 (4.94) dollars (Cox, 2003). Furthermore, our experiments confirm the finding that only a small fraction of players conform to the subgame perfect equilibrium with pure selfishness by giving nothing (10 percent of first movers and 22 percent of second movers across the three sessions). In line with most of the existing comparable experimental evidence, our results therefore appear incompatible with universal selfishness as the sole, or even dominant, motivation in trust-game settings.[22]

5.1 Is It Altruism?

First, we estimate equation (5), a simple regression of s_i on y_j. The results are given in column (I) of Table 2. We find a coefficient on y_j of 0.02 which has the "wrong" sign and is statistically insignificant. Virtually the same result obtains when we restrict the estimation to the "within" design of Session B and we control for subject-specific fixed effects (column (II)): y_j does not significantly affect s_i even in this most propitious of experimental designs. The null hypothesis of no altruism cannot therefore be rejected.

Next, we estimate a multi-group version of equation (6) by controlling for group-specific attributes that might affect mean transfers.[23] We consider five attributes: gender (*Female* = 1 for women), nationality (*Nat_Swiss* = 1 for Swiss nationals or permanent residents), native tongue (*Lang_French* = 1 for French speakers, *Lang_German* = 1 for German speakers), subject of study (*Non-economist* = 1 for non-economics/business students) and experimental session (*Session_B* = 1 for Session B; *Session_C* = 1 for Session C). Table 1 reports summary statistics on the distribution of those attributes in our subject

---

what we aim for, this design appears particularly attractive to our purpose, and we compare results of within- and between-subjects protocols.

[22] Our somewhat guarded use of language here is due to the fact that, in principle, non-zero transfer levels could be due to experimental bias.

[23] Note that the RESET test for model (I) in Table 2 indicates no misspecification problem. Hence, our parsimonious model, although almost devoid of explanatory power, does not appear fraught with estimation bias

sample. Estimation results are given in column (III) of Table 2. The coefficient on y_j is substantially unaffected, and the altruism hypothesis is therefore again not supported. Gender, nationality and mother tongue have no statistically significant impact on first-mover transfers either.[24] We find, however, that non-economics students send significantly more than economics and business majors.[25] There is also a borderline-significant effect of experimental Session B - a finding which highlights the potential biases affecting between-treatment comparisons.

In a third step, we extend the multi-group specification to allow also for different altruism according to group attributes, by adding interaction effects as in equation (7). The last column of Table 2 reports our estimations. We now find that the coefficient on y_j has the "correct" negative sign, but it continues to be statistically insignificant. More importantly, we find none of the interaction effects to be statistically significant. It is particularly revealing that not even the interaction term for Session B ($Session\_B \times y_j$) is significant, recalling that in that session a within-subject protocol was adopted and any impact of y_j on s_i could be expected to be particularly strong there. Furthermore, *F* tests on the joint significance of interactions with all group attributes or subsets thereof all fail to reject the null hypothesis that the coefficients are jointly zero.

To account for the two-sided censoring of s_i implied by the trust game, we re-estimated the four equations using the two-sided Tobit estimator (Table 3). The results are qualitatively unchanged from

---

due to omitted variables. That is of course not surprising, given that the randomized design of the experiment should make y_j, the sole regressor of model (I), uncorrelated with any player characteristics.

[24] Our results mirror those of Glaeser *et al.* (2000), who found that a range of similar control variables in a subject pool consisting of undergraduate students did not significantly affect s_i. This need not mean, however, that there are no group-specific differences in reciprocity or altruism. Playing dictator games with varying payoff structures, Andreoni and Vesterlund (2001), for example, observed that "demand curves for altruism" of men and women are different but cross at a certain "price of giving". One possible explanation for insignificant group effects could therefore be that the reward structure of our experiment is such that it places members of different groups close to those crossing points.

the OLS runs. The coefficient on y_j is never significant, in terms of both main effects and interaction terms.

In sum, our results suggest that altruism is *not* a statistically significant motivating force in determining "trust-like" behavior, both across all subjects and for specific groups of players.

5.2 Is It Reciprocity?

Table 4 reports regression estimates based on equation (8), in univariate form (column (I)) and with group-specific controls (column (II)). The univariate model yields an (implausibly) negative coefficient, but the borderline-significant *P* value on the RESET tests suggests misspecification bias. When we include controls for subject attributes and experimental treatments, the coefficient on $\hat{\rho}_j$ turns positive, as expected, but remains statistically insignificant. The magnitude of the coefficient implies that a first mover who expects the second mover to return 1.5 times s_i, earning him a 50% "profit", sends only 0.08 francs more than a first mover who expects the second mover merely to return s_i, which would leave him with no gain.

As discussed above, this analysis is biased against detecting trust, due to the fact that $\hat{\rho}_j$ is observed through questionnaire answers and thus likely measured with error. One particular issue with our protocol is that we asked first movers their $\hat{\rho}_j$ *after* they had observed ρ_j. This sequencing was motivated by the practical expediency of handing out questionnaires at the end of the game in manual sessions, but it implies a risk of biased *ex post* reporting. Figure 2 plots observed returns against first-

[25] This result also conforms to prior findings. In an overview of relevant empirical studies, Frank, Gilovich and Regan (1993, p. 170) conclude that there is "a large difference in the extent to which economists and noneconomists behave self-interestedly", and that "economists are more likely than others to free-ride".

movers' stated expectations. The scatter does not suggest strong positive correlation, and the sample correlation coefficient equals 0.33. Observation bias thus does not appear to be strong.

We can limit one distorting impact of mismeasurement by dropping observations for which inaccurate reporting appears particularly probable. In a first step, we drop all observations with $\hat{\rho}_j$ bigger than 3, which would suggest that first movers would have expected second movers to return more than the total transfer of $3s_i$ received and is thus incompatible with our assumed preference structure as well as with any other behavioral theory. Dropping the four observations concerned has no qualitative impact on control variables but changes the result on $\hat{\rho}_j$ dramatically (Table 4, columns (III) and (IV)). The estimated coefficient is now statistically significantly positive, which suggests that expected reciprocation is a significant determinant of first-mover transfers. We also find that the quantitative impact of expected reciprocity increases more than five-fold when we drop the four highly implausible observations, from 0.17 (column (II)) to 0.94 (column (IV)).

As another check on our results, we drop all observations with reported $\hat{\rho}_j$ exceeding 2. A $\hat{\rho}_j$ larger than 2 implies that second movers are expected to return more than two thirds of the tripled transfer, which would necessitate strong expected altruism on the part of the second mover. Given our lack of evidence of altruism in first-mover behavior, it seems implausible for first movers to expect strong altruism on the part of their paired second movers. The results, given in columns (V) and (VI) of Table 4, again suggest that expected second-mover returns significantly increase first-mover transfers. The estimated coefficient of the last specification suggests that a first mover who expects the second mover to return 1.5 times s_i, earning her a 50% "profit", sends 0.59 francs more than a first mover who expects the second mover merely to return s_i. Given the attenuation bias in our estimation, this must be considered as a lower-bound estimate.

We also estimated the model with $\hat{\rho}_j$ with the full set of interactions (analogously to equation (7)). None of the interaction terms was found to be statistically significant. This implies that there are neither

group-specific differences in reciprocity-motivated transfers, nor does the reciprocity motive appear with different strength in the within-subjects treatment compared to the between-subjects treatments. The latter result is of course not surprising, as our model does not predict that reciprocity-based trust should be affected by y_j. [26]

In sum, our results suggest that trust is a statistically significant motivating force in determining "trust-like" behavior. Trust-based first-mover giving seems to be based on a generally shared norm that does not differ significantly across subject groups.

6. CONCLUSIONS

We propose discriminatory criteria to identify altruism and reciprocity as determinants of first-mover transfers in trust games. The tests are based on within-treatment and, in one experimental session, on within-subject comparisons. They should therefore be immune to the experimental bias problem associated with the random component in the choices of laboratory subjects. Post-experimental questionnaires furthermore allow us to control for potential group-specific effects on trust-game transfers. Inference on our results, based on experiments using randomized double-blind protocols and conducted with University of Lausanne undergraduate students, reject altruism but accept expected reciprocity as explanations for "trust-like" transfers.

[26] We have experimented with more refined selection criteria, by setting differentiated plausibility thresholds for reported $\hat{\rho}_j$ according to second-mover endowments y_j (with maximum plausible $\hat{\rho}_j$ increasing in y_j), but the results remained qualitatively unchanged. In addition, we estimated the model using common methods for dealing with mismeasured regressors, including bootstrap estimation, inverse least squares and the method of grouping. All these approaches confirmed the statistically significantly positive coefficients on $\hat{\rho}_j$. We also reestimated the model (i) using the Tobit estimator and (ii) considering only the "within" variation of Session B, but we found the results qualitatively unchanged. Finally, to take account of non-normal disturbances, we estimated all models using the LAD estimator and bootstrap confidence intervals. Again, the results were not substantially changed. All these results are available from the authors on request.

Our findings lend support to the view that social preferences in extensive non-repeated games are not separable: perceived kindness and intentions matter. Related studies have come to similar conclusions, but from the point of view of *second movers*, i.e. from agents who base their choices on their interpretation of the "kindness" implied in first movers' observed decisions (Clark and Sefton, 2000; Charness and Haruvy, 2002; Cox, 2003; McCabe *et al.* 2003; Nelson, 2002). We confirm that *first movers*' choices are significantly determined by the anticipation of reciprocal behavior on the part of second movers: what looks like trust, seems to be trust. Trust games therefore do seem to be a valid method to fill the "great lacuna in this research agenda [that is] the measurement of trust" (Glaeser *et al*., 2000, p. 811).

Notwithstanding this positive result, a cautionary note is also in order. We have insisted on the simple point that "noise" should be taken into account when interpreting experimental results. The whole experimental approach to economics strives to eliminate as much as possible any non-controlled influences. The "as much as possible" qualifier is important: when even physicists cannot create 100% controlled laboratory conditions, economists must be realistic about their ability to eliminate unintended influences on subject behavior. This does not undermine the validity of economic experiments, but it calls for carefully designed hypothesis tests.

This research could be extended in a number of ways. It would be interesting for our experiment to be replicated with different subject pools, since, as pointed out e.g. by Fehr *et al*. (2003), one type of experimental bias could arise through the non-representativeness of self-selected student samples. The only way to correct for bias of this type is to make subject pools more representative. Another potentially worthwhile modification would be to use higher monetary stakes, to test whether our rejection of the altruism hypothesis is robust to a compression of the variance of the disturbance term. Finally, our model implies common altruism and reciprocity preferences across subjects of a particular type, and attributes any individual idiosyncrasy to the noise term. The within-subject experimental approach might be extended in a way to allow estimation of a model that accommodates individual-specific tastes for altruism and reciprocity.

BIBLIOGRAPHY

Abbink, Klaus; Irlenbusch, Bernd and Renner, Elke (2000) "The Moonlighting Game: An Experimental Study on Reciprocity and Retribution". *Journal of Economic Behavior and Organization*, 42: 265-277.

Alesina, Alberto and La Ferrara, Eliana (2002) "Who Trusts Others?" *Journal of Public Economics*, 85: 207-234.

Andreoni, James (1995a) "Warm-Glow Versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments". *Quarterly Journal of Economics*, 60(1): 1-21.

Andreoni, James (1995b) "Cooperation in Public-Goods Experiments: Kindness or Confusion". *American Economic Review*, 85(4): 891-904.

Andreoni, James and Miller, John (2002) "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism". *Econometrica*, 70(2): 737-753.

Andreoni, James and Vesterlund, Lise (2001) "Which is the Fair Sex? Gender Differences in Altruism". *Quarterly Journal of Economics*, 116: 293-312.

Arrow, Kenneth J. (1974a) *The Limits of Organization*. W.W. Norton, New York.

Arrow, Kenneth J. (1974b) *Essays in the Theory of Risk Bearing*, North-Holland, Amsterdam.

Ben-Ner, Avner; Putterman, Louis; Kong, Fanmin and Magan, Dan (2004) "Reciprocity in a Two-Part Dictator Game". *Journal of Economic Behavior and Organization*, 53: 333-352.

Berg, Joyce; Dickhaut, John and McCabe, Kevin (1995) "Trust, Reciprocity and Social History", *Games and Economic Behavior*, 10: 122-142.

Bohnet, Iris and Frey, Bruno S. (1999) "Social Distance and Other-Regarding Behavior in Dictator Games: Comment". *American Economic Review*, 89:335-341.

Bolton, Gary E. and Ockenfels, Axel (2000) "ERC: A Theory of Equity, Reciprocity and Competition". *American Economic Review*, 90: 166-193.

Bolton, Gary E.; Katok, Elena and Zwick, Rami (1998) "Dictator Game Giving: Rules of Fairness Versus Acts of Kindness". *International Journal of Game Theory*, 27: 269-299.

Bornhorst, Fabian; Ichino, Andrea; Schlag, Karl and Winter, Eyal (2004) Trust and Trustworthiness among Europeans: South-North Comparison. *CEPR Discussion Paper*, No. 4378.

Buchan, Nancy R.; Croson, Rachel T.A. and Dawes, Robyn M. (2001) "Direct and Indirect Trust and Reciprocity". *Mimeo*, Wharton School, University of Pennsylvania.

Burnham, Terence; McCabe, Kevin and Smith, Vernon L. (2000) "Friend-or-Foe Intentionality Priming in an Extensive Form Trust Game". *Journal of Economic Behavior and Organization*, 43: 57-73.

Camerer, Colin F. (2003) *Behavioral Game Theory*. Princeton University Press.

Camerer, Colin F. and Hogarth, Robin M. (1999) "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework". *Journal of Risk and Uncertainty*, 19: 7-42.

Charness, Gary (2004) "Attribution and Reciprocity in an Experimental Labor Market". *Journal of Labor Economics*, forthcoming.

Charness, Gary; Frechette, Guillaume R. and Kagel, John H. (2004) "How Robust is Laboratory Gift Exchange?". *Experimental Economics*, 7: 189-205.

Charness, Gary and Haruvy, Ernan (2002) "Altruism, Equity, and Reciprocity in a Gift-Exchange Experiment: An Encompassing Approach". *Games and Economic Behavior*, 40: 203-231.

Clark, Kenneth and Sefton, Martin (2001) "The Sequential Prisoner's Dilemma: Evidence on Reciprocation". *Economic Journal*, 111: 51-68.

Cox, James C. (2001) "On the Economics of Reciprocity". *Mimeo*, University of Arizona.

Cox, James C. (2003) "Trust and Reciprocity: Implications of Game Triads and Social Contexts". *Mimeo*, University of Arizona.

Cox, James C. (2004) "How to Identify Trust and Reciprocity". *Games and Economic Behavior*, 46: 260-281.

Cox, James C. and Friedman, Daniel (2002) "A Tractable Model of Reciprocity and Fairness". *Mimeo*, University of Arizona.

Cox, James C.; Sadiraj, Klarita and Sadiraj, Vjollca (2002) "Trust, Fear, Reciprocity, and Altruism". *Mimeo*, University of Arizona.

Dufwenberg, Martin and Gneezy, Uri (2000) “Measuring Beliefs in an Experimental Lost Wallet Game”. *Games and Economic Behavior*, 30: 163-182.

Eckel, Catherine C. and Grossman, Philip J. (1996) “Altruism in Anonymous Dictator Games”. *Games and Economic Behavior*, 16: 181-191.

Fehr, Ernst and Fischbacher, Urs (2002) “Why Social Preferences Matter - The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives”. *Economic Journal*, 112: C1-C33.

Fehr, Ernst; Fischbacher, Urs; von Rosenbladt, Bernhard; Schupp, Jürgen and Wagner, Gert G. (2003) “A Nation-Wide Laboratory: Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Surveys”. *CESifo Working Paper*, No. 866, Munich.

Fehr, Ernst and Gächter, Simon (2000) “Fairness and Retaliation: The Economics of Reciprocity”. *Journal of Economic Perspectives*, 14: 159-181.

Fehr, Ernst; Kirchsteiger, Georg and Riedl, Arno (1993) “Does Fairness Prevent Market Clearing? An Experimental Investigation”. *Quarterly Journal of Economics*, 108: 437-460.

Fehr, Ernst and List, John A. (2004) “The Hidden Costs and Returns of Incentives - Trust and Trustworthiness among CEOs”. *Journal of the European Economic Association*, 2: 743-771.

Fehr, Ernst and Schmidt, Klaus (1999) “A Theory of Fairness, Competition and Cooperation”. *Quarterly Journal of Economics*, 114: 817-868.

Fershtman, Chaim and Gneezy, Uri (2001) “Discrimination in a Segmented Society”. *Quarterly Journal of Economics*, 116: 351-377.

Frank, Robert H.; Gilovich, Thomas and Regan, Dennis T. (1993) “Does Studying Economics Inhibit Cooperation?”. *Journal of Economic Perspectives*, 7(2): 159-171.

Glaeser, Edward L.; Laibson, David; Scheinkman, Jose A. and Soutter, Christine L. (2000) “Measuring Trust”. *Quarterly Journal of Economics*, 115: 811-846.

Guiso, Luigi; Sapienza, Paola and Zingales, Luigi (2004) “Cultural Biases in Economic Exchange”. *Mimeo*, Northwestern University and University of Chicago.

Hoffman, Elizabeth; McCabe, Kevin and Smith, Vernon L. (1996) “Social Distance and Other-Regarding Behavior in Dictator Games”. *American Economic Review*, 86: 653-660.

Houser, Daniel and Kurzban, Robert (2002) “Revisiting Kindness and Confusion in Public Goods Experiments”. *American Economic Review*, 92(4): 1062-1069.

Jenks, Christopher (1990) “Varieties of Altruism”. In: Mansbridge, Jane J. (Ed.) *Beyond Self-Interest*, University of Chicago Press, 53-70.

La Porta, Rafael; Lopez-de-Silanes, Florencio; Shleifer, Andrei and Vishny, Robert W. (1997) “Trust in Large Organizations”. *American Economic Review*, 87: 333-338.

McCabe, Kevin A.; Rigdon, Mary L. and Smith, Vernon L. (2003) “Positive Reciprocity and Intentions in Trust Games”. *Journal of Economic Behavior and Organization*. 52: 267-275.

Meijer, Erik and Wansbeek, Tom (2000) “Measurement Error in a Single Regressor”. *Economics Letters*, 69: 277-284.

Nelson, William R. (2002) “Equity or Intention: It is the Thought that Counts”. *Journal of Economic Behavior and Organization*, 48: 423-430.

Ortmann, Andreas; Fitzgerald, John and Boeing, Carl (2000) “Trust, Reciprocity, and Social History: A Re-examination”. *Experimental Economics*, 3: 81-100.

Rabin, Matthew (1993) “Incorporating Fairness into Game Theory and Economics”. *American Economic Review*, 83: 1281-1302.

Smith, Vernon L. (2003) “Constructivist and Ecological Rationality in Economics”. *American Economic Review*, 93(3): 465-508.

Sobel, Joel (2004) “Interdependent Preferences and Reciprocity”. *Mimeo*, University of California, San Diego.

Willinger, Marc; Keser, Claudia; Lohmann, Christopher and Usunier, Jean-Claude (2003) “A Comparison of Trust and Reciprocity between France and Germany: Experimental Investigation Based on the Investment Game”. *Journal of Economic Psychology*, 24: 447-466.

Zack, Paul J. and Knack, Stephen (2001) “Trust and Growth”. *Economic Journal*, 111: 295-321.

Table 1: Data Description

	Session A	**Session B**	**Session C**	**TOTAL**
No. of observations	38	36[#]	31	105
s_i [##]	7.76 (2.63)	6.44 (3.49)	6.77 (4.31)	7.02 (3.50)
r_j [##]	12.37 (10.93)	8.06 (8.19)	10.45 (13.93)	10.32 (11.15)
y_i [##]	10 (0)	10 (0)	10 (0)	10 (0)
$\hat{\rho}_j$ [##]	2.00 (1.12)	1.86 (0.95)	1.32 (0.82)	1.78 (1.01)
y_j [###¶]	13 * 0 12 * 10 13 *20	18 * 0 18 * 20	16 * 0 15 * 20	47 * 0 12 * 10 46 * 20
Female	18.4%	38.9%	19.4%	25.7%
Nat_Swiss	92.1%	83.3%	83.9%	86.7%
Lang_French	81.6%	77.8%	77.4%	79.0%
Lang_German	13.2%	11.1%	9.7%	11.4%
Non-economist	2.6%	11.1%	0.0%	5.8%

[#] In Session 2, 36 observations correspond to 18 players 1, each matched with two players 2.
[##] Mean values (standard deviations in parentheses)
[###] (number of observations × y)

Table 2: Altruism Regressions, OLS
(independent variable = s_i) [#]

	(I)	(II) [##]	(III)	(IV) [###]
y_j	0.02 (0.51)	0.02 (0.33)	0.02 (0.46)	-0.27 (-0.22)
Female			-0.62 (-0.76)	-0.27 (-0.22)
Female × y_j				-0.03 (-0.30)
Nat_Swiss			0.12 (0.11)	-0.64 (-0.45)
Nat_Swiss × y_j				0.08 (0.66)
Lang_French			-0.07 (-0.06)	1.05 (0.54)
Lang_French × y_j				-0.13 (-1.00)
Lang_German			1.34 (0.85)	3.28 (1.48)
Lang_German × y_j				-0.19 (-1.23)
Non-economist			3.66 (4.48) ***	3.34 (2.72) ***
Non-economist × y_j				0.04 (0.43)
Session_B			-1.47 (-1.88) *	-1.92 (-1.55)
Session_B × y_j				0.04 (0.38)
Session_C			-0.83 (-0.90)	-1.42 (-1.02)
Session_C × y_j				0.06 (0.56)
Dummies for 1st-movers	no	yes	no	no
Constant	6.83 (13.71) ***	4.83 (3.71) ***	8.73 (5.45) ***	10.56 (5.45) ***
R-squared	0.003	0.64	0.09	0.11
F statistic (full model)	0.27	133.34 ***	3.56 ***	3.70 ***
Breusch-Pagan test [####]	0.91	0.005	0.11	0.07
RESET test [#####]	0.23	0.94	0.92	0.64
Observations	105	36	105	105

[#] White-corrected *t* statistics in brackets; *: 90% confidence level, **: 95% , ***: 99%
[##] Regression includes only observations from Session B.
[###] *F* statistic (*P* value) on H_0 that interaction terms are jointly zero: all 7 variables: 0.42 (0.89), 2 language variables: 0.78 (0.46), 2 session variables: 0.17 (0.84)
[####] *P* value of chi-square test of constant error variance
[#####] *P* value of *F* test of statistical significance of powers of fitted values (H_0: correct functional form, no omitted variables)

Table 3: Altruism Regressions, Tobit
(independent variable = s_i) [#]

	(I)	(II) [##]	(III)	(IV) [###]
y_j	0.06 (0.61)	0.02 (0.36)	0.05 (0.57)	-0.38 (-0.77)
Female			-1.68 (-0.91)	-0.92 (-0.38)
Female × y_j				-0.07 (-0.34)
Nat_Swiss			0.46 (0.18)	-1.37 (-0.38)
Nat_Swiss × y_j				0.18 (0.67)
Lang_French			-0.12 (-0.04)	2.15 (0.47)
Lang_French × y_j				-0.27 (-0.68)
Lang_German			3.80 (0.93)	8.60 (1.39)
Lang_German × y_j				-0.45 (-1.09)
Non-economist			2.01 (2.03) **	6.27 (1.27)
Non-economist × y_j				2.03 (n.a.) [####]
Session_B			-2.75 (-1.56)	-3.71 (-1.38)
Session_B × y_j				0.06 (0.30)
Session_C			-1.10 (-0.50)	-3.01 (-0.98)
Session_C × y_j				0.19 (0.80)
Dummies for 1st-movers	no	yes	no	no
Constant	8.93 (7.21)***	4.76 (1.96) *	13.34 (3.42)***	18.48 (5.45) **
Pseudo R-squared [#####]	0.001	0.28	0.02	0.03
Observations	105	36	105	105

[#] two-sided Tobit with censoring at s_i =0 and s_i = 10; White-corrected *t* statistics in brackets;
*: 90% confidence level, **: 95%, ***: 99%
[##] Regression includes only observations from Session B.
[###] *F* statistic (*P* value) on H_0 that interaction terms are jointly zero: all 7 variables: 0.33 (0.92), 2 language variables: 0.43 (0.65), 2 session variables: 0.42 (0.66)
[####] not estimated due to insufficient degrees of freedom
[#####] = 1 - L_1/L_0, where L_0 and L_1 are the constant-only and full model log-likelihoods respectively

Table 4: Reciprocity Regressions, OLS
(independent variable = s_i) [#]

Observations included if:	(I)	(II)	(III)	(IV)	(V)	(VI)
	$s_i > 0$		$s_i > 0,\ \hat{\rho}_j \leq 3$		$s_i > 0,\ \hat{\rho}_j \leq 2$	
$\hat{\rho}_j$ [##]	-0.02 (-0.06)	0.17 (0.47)	0.72 (1.79) **	0.94 (2.44) ***	0.90 (1.34) *	1.17 (1.77) **
y_j		0.01 (0.37)		0.01 (0.31)		0.01 (0.34)
Female		-1.12 (-1.73) *		-1.20 (-1.93) *		-0.88 (-1.22)
Nat_Swiss		0.79 (0.94)		0.43 (0.51)		-0.13 (-0.15)
Lang_French		-0.37 (-0.41)		-0.22 (-0.23)		0.01 (0.01)
Lang_German		1.21 (1.18)		1.37 (1.30)		1.34 (1.16)
Non-economist		2.93 (4.17) ***		2.73 (3.99) ***		2.96 (3.79) ***
Session_B		-0.65 (-0.94)		-0.33 (-0.52)		-0.50 (-0.61)
Session_C		1.14 (1.68) *		1.43 (2.21) **		1.45 (1.70) *
Constant	7.95 (12.38) ***	7.05 (6.53) ***	6.86 (9.02) ***	5.93 (5.46) ***	6.68 (6.60) ***	5.92 (4.08) ***
Observations [###]	93	93	89	89	69	69
R-squared	0.0001	0.18	0.04	0.23	0.04	0.11
F statistic (full model)	0.00	3.75 ***	3.20 *	3.72 ***	1.78	2.80 ***
Breusch-Pagan test [####]	0.95	0.19	0.09	0.06	0.31	0.31
RESET test [#####]	0.12	0.82	0.31	0.57	0.001	0.77

[#] White-corrected *t* statistics in brackets; *: 90% confidence level, **: 95% confidence level, ***: 99% confidence level

[##] confidence levels with respect to one-tailed *t* test

[###] 11 observations with $s_i = 0$ and one observation with unreported $\hat{r}_i$ were dropped.

[####] *P* value of chi-square test of constant error variance

[#####] *P* value of *F* test of statistical significance of powers of fitted values (H_0: correct functional form, no omitted variables)

Figure 1: First-Mover Transfers and Second-Mover Returns

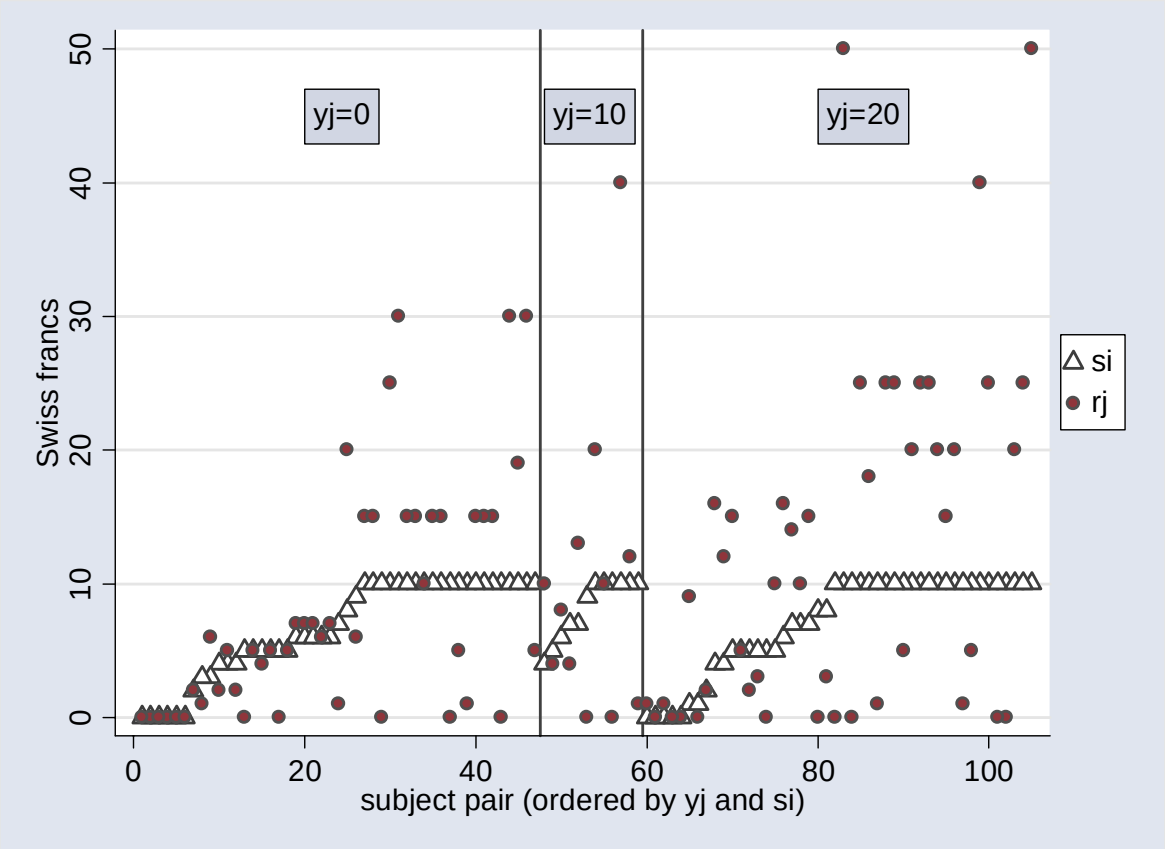

Figure 2: Expected and Actual Second-Mover Returns