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

MARIUS BRÜLHART
University of Lausanne & CEPR

JEAN-CLAUDE USUNIER
University of Lausanne

May 2008

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 laboratory "trust games". However, recent research suggests that altruism may explain much of what "looks like" trust in such experiments. We formally derive discriminatory tests for altruism and trust based on within-treatment and within-subject comparisons, and we control for group attributes of experimental subjects. The central idea is to allow for rich and poor trustees, and to examine whether, consistent with dominant altruism, trustors give more to the poor, or whether, consistent with dominant reciprocity motives, trustors give no more to the poor than to the rich. Our results support trust as the dominant motivation for "trust like" decisions, with at most a subsidiary role for altruism.

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 on an earlier version of this paper from Olivier Cadot, Simon Gächter, Ulrich Hoffrage and Mario Jametti as well as from participants at the 2005 Econometric Society World Congress (University College London), the 2006 Annual Conference of the Royal Economic Society (Nottingham) and at seminars at McMaster and Nottingham Universities. Hansueli Bacher, Gregory Cleusix, Tea Danelutti, Jerôme Lathion 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 RANDOMNESS 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 equality or "fairness". This basic theoretical building block has recently 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*.[1]

We focus on the measurement of trust.[2] The "trust game" of Berg, Dickhaut and McCabe (1995) has become a standard experiment to measure trust.[3] In the trust game, a *first mover* is randomly and

[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; Fehr, Klein and Schmidt, 2007), 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 international investment and trade (Guiso, Sapienza and Zingales, 2007) and to increase economic growth (Zak and Knack, 2001). For a critical appraisal, see Durlauf (2002).

[2] 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, Friedman and Gjerstad, 2007; 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.

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. First-mover transfers are interpreted as a manifestation of trust, and second-mover transfers as a manifestation of trustworthiness.[4]

Our central idea is a simple twist to this game: we produce "rich" and "poor" second movers by giving them different experimental endowments, and we observe how this distinction affects first-mover transfers, controlling for expected reciprocation. We propose this amended protocol as a way to investigate the claim that *altruism* might play a significant role in "trust like" first-mover transfers, and we derive our discriminatory criteria formally from an explicit model of agents' preferences.[5]

Furthermore, we argue that the interpretation of observed trust-game transfers is complicated by *randomness*. There may be idiosyncrasies in individual preferences and, more importantly, potential biases induced by the structure, framing and practical implementation of experiments. The mere fact that a game is based on a single decision node (such as the dictator game) or on a sequence of nodes (such as

---

[3] The trust game is sometimes also referred to as the "investment game". The seminal experiment of this kind is the "gift-exchange game" of Fehr, Kirchsteiger and Riedl (1993). The main differences between the (Fehr *et al*., 1993, variant of 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 proposed 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.

[4] 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, Laibson, Scheinkman and Soutter (2000) observe only weak correlations 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 for a broad cross-section of German residents, find that survey responses have strong predictive power for first-mover trust-game transfers; and Gächter, Herrmann and Thöni (2004) observe a strong correlation between subjects' stated trust attitude and their behavior in a laboratory public-good experiment.

[5] We use the "pure" definition of altruism (Andreoni, 1990), implying a concern for other individuals' payoffs *per se* (see Section 2.2).

the trust game) may influence transfers, and difficult-to-control details of practical implementation can introduce treatment-specific biases. By differentiating second movers by their experimental endowment, we can run our discriminatory test *within treatments* and even *within subjects* (when we let each first mover play simultaneously with a poor and a rich second mover). Thus, we can control for potential individual-specific biases as well as for treatment-specific bias.

Using data gathered in four experimental sessions with undergraduate university students, we do not find evidence of a significant negative relation between first-mover transfers and second-mover "wealth". This result rejects the hypothesis of altruistic motives as the dominant determinant of "trust-like" decisions.

Another feature of our study is to elicit subjects' expectations in order to inform our identification of motivations. Answers to a question on first movers' expected returns allow us to test the original interpretation of observed trust-game decisions, that is whether "trust-like" transfers are indeed significantly affected by expected reciprocation. Even though elicited expectations may be prone to measurement error, and estimated coefficients on those measured expectations could thus be biased toward zero, we find significantly positive coefficients when regressing first-mover transfers on expected second-mover returns.[6] This result supports the hypothesis that "trust-like" behavior is indeed motivated mainly by trust, i.e. first movers' expectation of reciprocal kindness on the part of second movers. In addition, we find evidence that, once we control for reciprocity-based motivations, altruistic motives play a subsidiary role.

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

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

described in Section 4. Section 5 reports our experimental findings and tests the two discriminating hypotheses econometrically. Section 6 concludes.

2. PLAYING GAMES WITH RECIPROCITY AND ALTRUISM: 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 about the individual concerned. In contrast, altruism is often considered to be "context dependent" (Eckel and Grossman, 1996), meaning that it is negatively related to social distance (Jencks, 1990; Bohnet and Frey, 1999).

In the double-blind one-shot trust game experiment, 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 trustworthiness. Agents in such games have consistently made significantly positive transfers in both directions.

Observed transfers in this game have been interpreted as strong evidence for the pervasiveness of trust and trustworthiness as a motivators of social behavior, and they allow for interesting intercultural comparisons.[7] In the United States, Berg *et al*. (1995) observed that first movers on average entrusted 52% of their endowment to second movers. Replicating the 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. Holm and Danielson (2005) found Tanzanian and Swedish students to be similarly trusting, as they on average transferred 53% and 51% of their endowment respectively. Similarly, Buchan, Croson and Johnson (2006) observed only small differences in trust among Chinese, Korean, Japanese and US subjects. Fershtman and Gneezy (2001) concluded that Israeli subjects considered Ashkenazic second movers more trustworthy than Sephardic second movers, since average first-mover transfers to the respective recipient types corresponded to 76% and 40% of first-movers' endowments. 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%.

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

Altruism has many definitions.[8] We understand it to mean a concern for other individuals' payoffs *per se*. This can for instance take the form of inequality aversion (e.g. Fehr and Schmitt, 1999) or quasi-maximin preferences (e.g. Charness and Rabin, 2002). Hence, we use the term altruism in its "pure" variant (Andreoni, 1990) that does not include "warm glow" motivations for doing good. Warm glow effects are individual-specific and unrelated to relative payoffs (Palfrey and Prisbrey, 1997), which is why we treat them as a random preference component (see Section 2.3).

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

[8] The literature abounds with terminology for utility functions that include arguments other than own payoff (equity, fairness, other-regarding preferences, social preferences, etc.). In most cases, these terms are used in a way that encompasses altruism and reciprocity (e.g. Buchan *et al*., 2006). One motivation for our formalizations in Section 3 is to state precisely the meaning we place on the terms we use.

In order to interpret non-zero transfers in standard trust games as measures of reciprocity, one has to assume away warm-glow effects. Furthermore, "pure" altruistic motives cannot be allowed to 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.[9]

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 the (overwhelmingly) *dominant* determinant of transfers in trust games. However, it cannot rule out

[9] 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

altruism as *part of* the motivation underlying "trust-like" giving (altruism just is not strong enough for first movers to make the sacrifice implied in the particular altruistic allocation that is on offer). Moreover, opting for the altruistic allocation in this game is in fact incompatible with altruism defined as inequality aversion, 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.[10] Members of the first group of dictators are given amounts that are equal to those allocated to first movers in the trust 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% of first-mover trust-game transfers. Ashraf, Bohnet and Piankov (2006) run the triadic experiment using a within-subject design, where the same subject participates in the dictator and the trust game. Their subjects, students in Russia, South Africa and the United States, on average sent 25% and 45% respectively of their endowment when playing the trust game and the dictator game. Hence, mean dictator-game transfers amount to 55% of mean trust-game transfers.[11] In a comparable within-subject triadic game conducted in South Africa, Carter and Castillo (2006) even find that mean dictator transfers

(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.

[10] Cox, Sadiraj and Sadiraj (2007) apply the triadic setup to the moonlighting game.

[11] These percentages can be found in the working-paper version of their paper.

amounted to a full 81% of corresponding trust-game transfers. These results suggest that a major share of what has commonly been interpreted as trust-based transfers may in fact be motivated by altruism.

Similar evidence is found in several other studies. Buchan, Croson and Dawes (2002) 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 61% of transfers in the standard trust game. Dufwenberg and Gneezy (2000) compare dictator transfers to second-mover transfers made in an experiment that 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.[12] They observe that while 65% of second movers transfer five in the trust-game treatment, a full 33% of second movers transfer five in the dictator-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".

[12] The McCabe *et al*. (2003) experiment in fact pre-dated Cox's studies using the "triadic" setup.

2.3 A Further Complication: Randomness

In addition to selfishness, reciprocity 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. More seriously still, randomness could imply *experimental bias*, if motivational features that are beyond the control of the experimenter lead to systematic divergence between subjects' decisions and their hypothetical preferred choice under perfect information, sufficient stakes and neutral framing (see, e.g., Levitt and List, 2007).

One source of randomness is incomplete information-processing by experimental subjects. 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).

We consider "warm glow" altruism as another source of randomness. Warm glow is triggered by the very act of giving, irrespective of final payoffs and of other players' observed or expected actions (Andreoni, 1990). Individuals' warm-glow attitudes are likely to be heterogeneous and responsive to the structure and presentation of experiments (Palfrey and Prisbrey, 1997).

Given the plausibility of confusion and warm-glow effects, it is not surprising 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 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 toward 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 (pure) 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.* (2002) 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.[13] The within-subject design of Ashraf *et al.* (2006) avoids this potential complication. However, more fundamentally still, different behavioral norms could

[13] 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." Levitt and List (2007) forcefully argue in favor of experimental protocols that allow the researcher to "net out" laboratory effects.

be triggered by varying experimental protocols, even for within-subject designs. 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 simple curiosity about what will happen to them when the game has more than one decision node. This is relevant for the "triadic" game, since curiosity about later stages may play a role in the trust game but not in the dictator game.

For valid statistical inference, one would therefore wish to make comparisons within-session or, even better, within-subject, rather than between-session. Furthermore, experimental protocols should be held as similar as possible. In such settings, experimental biases affect treatments and controls identically and thus wash out in the comparison.

3. ALTRUISM VERSUS TRUST VERSUS RANDOMNESS: 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.[14]

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.[15] 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

[14] Levitt and List (2007, p. 170) argue in favour of "combining laboratory analysis with a model of decision-making". Our behavioral model can be seen as a specific variant of the general utility function they propose.

[15] 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.[16] 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,

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

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

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

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

[16] In the Berg *et al*. (1995) trust game, second mover were not allowed to use their initial money holding y_j as part of their transfer r_j. We relax this constraint, analogously to Buchan *et al*. (2006), by allowing r_j to include y_j.

Assumptions (1b) represent standard concave preferences over own income with nonincreasing absolute risk aversion.[17] Assumptions (1c) define altruism, incorporating an element of inequality aversion: my utility gain from a given increase in your payoff is larger if this leaves you poorer than me than if this makes you even richer than me.[18] Assumption (1d) represents "intrinsic reciprocity" (Sobel, 2005): 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.[19] Finally, (1e) expresses that agents' utility may be affected by idiosyncratic factors such as confusion and warm-glow sentiments.

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

- **(1f)**: First movers' expected rate of return ($\hat{\rho}_j$) and the variance thereof are independent of the amount sent (s_i).

Greig and Bohnet (2005) refer to this as "balanced" expected reciprocation. In the Appendix, we show that this assumption is consistent with our experimental data.

---

[17] (1b) implies that subjects do not regard the experimental payoffs simply as a (tiny) fraction of their lifetime wealth, as this would be incompatible with risk aversion (Rabin, 2000), but rather as a set of payoffs over which they hold preferences that are to some extent independent from their preferences over lifetime wealth levels (see, e.g., Samuelson, 2005).

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

[19] Sapienza, Toldra and Zingales (2007) refer to these two components of trust as "belief based" (the selfish component) and "preference based" (the intrisic reciprocity component).

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

- **(1g)**: 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:

- **(1h)**: f is additive in its arguments.
- **(1i)**: 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 (1i) because confusion and framing effects may bias s_i upward.[20]

Utility Function of Laboratory "Trustors"

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

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

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

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

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 (1g). The first three 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

Utility function (2) distinguishes four parameters determining first-mover transfers in the trust game (*a*, *b*, *c* and λ), but we only observe one variable s_i. The challenge is to design the experiment in such a way as to allow for the identification of separate determinants of s_i.

Our first approach is to vary second movers' initial wealth y_j, and to examine the relationship between s_i and y_j. Given the double-blind experimental protocol, y_j constitutes the only element of information that first movers have about their paired second movers. Consistency with their own utility function (2) implies that first movers expect richer second movers to return no less than poorer second movers

($\frac{d\hat{\rho}_j}{dy_j} \geq 0$), as richer second movers being able to "afford" a higher $\hat{\rho}_j$.[21] This, together with (1b-h) and (2), implies that:

$$\frac{\partial s_i}{\partial y_j} < 0 \quad \text{if} \qquad b > 0,\ a = 0,\ c = 0. \qquad \textbf{(3a)}$$

$$\frac{\partial s_i}{\partial y_j} \geq 0 \quad \text{if} \qquad b = 0,\ a \geq 0,\ c \geq 0. \qquad \textbf{(3b)}$$

(3a) states that if, except for randomness, altruism is the *sole* feature of first-movers' utility function, they will give more to poor second movers than to rich ones.[22] Conversely, according to (3b), first-mover transfers in the absence of own altruistic motives will non-negatively related to second-mover wealth. (3b) represents the behavioral model implicit in the original interpretation of trust-game results.

Expressions (3) thus provide a crisp discriminating hypothesis.

Proposition 1:

First movers who are motivated only by altruism send more to poor second movers than to rich second movers. First movers who are not motivated by altruism at all either send less to poor second movers than to rich second movers, or they send equal amounts to both.

Proposition 1 has the advantage of being based entirely on directly observable variables. It has the drawback, however, that it is not an explicit test of the relevance of *trust* as a motivating force underlying first-mover transfers. Another limitation is that it pits two extreme alternatives against each

[21] To be precise, this is true as long as $\hat{\rho}_j \leq 3$, since otherwise first-mover transfers would *reduce* second movers' final payoff.

[22] The qualitative result that s_i falls in y_j would hold even if altruism were linear, i.e. if the last derivative in (1c) 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.

other: *only* altruism versus *only* trust (where trust is a combination of selfishness and intrinsic reciprocity). To be empirically plausible, our behavioral model should be able to accommodate altruism and trust simultaneously, while providing us with a means of distinguishing the two.

If we extend the analysis to expected reciprocation $\hat{\rho}_j$, a variable that can be elicited from experimental subjects, we can postulate the following:

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

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

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.

As (4a) shows, trust has two components: intrinsic reciprocity (c) and selfish own-payoff maximization (a). 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.[23] 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 ($\hat{\rho}_j$).[24]

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 (1i), 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 may be problematic.

Our discriminating criterion can be evaluated through a t test on the null hypothesis that $\beta_{\text{OLS}} \geq 0$. 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 randomness in our data, altruism is not a significant determinant of s_i, and s_i thus represents a combination of trust and experimental error.

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,

---

[23] Bohnet and Zeckhauser (2004) show that trustors exhibit "exploitation aversion". This translates in our framework to $\hat{\rho}_j < 1$ combined with $c \neq 0$.

[24] Eckel and Wilson (2004), combining an amended trust game with measures of individual attitudes to risk, find little evidence that first-mover transfers are thought of as risky gambles. This evidence supports the interpretation of first-mover transfers in the anonymous one-shot game as motivated by an intrinsic taste for reciprocity.

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.[25]

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:

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

where β_1 captures the differential effect on altruism-induced transfers of female compared to male players.[26]

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.

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

3.4 Testing for Trust

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* versus an alternative of no trust. The trust-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 trust 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 motive. If both the null and the alternative hypotheses are rejected, i.e. $\phi_{\text{OLS}} < 0$, 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), as it can only be observed by asking subjects.[27] Model (8) therefore mixes experimental with survey methods. Given that the former have 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, ϕ_{OLS} and its associated standard error will be unambiguously biased toward zero (see, e.g., Meijer and Wansbeek, 2000). Since we have no perfect palliative for this

[26] When, as in our example, there is only one grouping variable, then separate regressions for the each group would be equivalent to estimating equation (7). When we control for multiple overlapping groupings, however, the interaction specification *à la* equation (7) is different from, and superior to, group-wise regressions.

[27] Note that $\hat{\rho}_j$ and y_j are uncorrelated in our formal behavioral model, which leaves β_{OLS} of regression equations (6) to (8) unbiased even though $\hat{\rho}_j$ is omitted from the estimations. Our result reported in Appendix Table 1, however, suggests that $\hat{\rho}_j$ in fact increases in y_j. Therefore, we explicitly examine our regressions for omitted-variables bias using the RESET test.

problem, our test for trust is biased in favor of acceptance of the null hypothesis of no trust. 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 per second mover they were paired with.[28] 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 four 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.[29]

- **Session A** was played manually with physical money (coins of 1 franc). 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). 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. 31 first movers played with one second mover each. 16 second movers started the game with nothing, and 15 started with 20 francs.

[28] One Swiss franc was worth approximately 0.73 and 0.85 US dollars in early 2003 and in early 2005 respectively.

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

- **Session D** was played via internet. 32 first movers played two different second movers each, on with no show-up fee, and on with a show-up fee of 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 that did not compromise their anonymity.[30]

Sessions C and D were 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 experimenters. Subjects were recruited via email and retained if they had not taken part in any of the previous sessions. 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. In Session D, first movers were asked to complete the questionnaire at the same time as making their decision. 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. In Session C, first movers were then asked to fill out a post-experiment questionnaire. Finally, all players 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.

[30] 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.

The manual sessions required about two hours, whereas the computerized sessions took between two and 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 have conducted the experiment in a variety of settings as a robustness check of our inference with respect to the experimental protocol used. This allows us to control for protocol-specific effects, and thus to curtail the potential bias that exists when observed transfers are compared across different sessions and/or protocols. Our approach is to allow variation in three methodological dimensions:

1. manual game (Sessions A and B) versus computer-based game (Sessions C and D),
2. between-subject design (Sessions A and C) versus within-subject design (Sessions B and D),[31] and
3. statement of expected returns *ex post* (Sessions A, B and C) versus statement of expected returns *ex ante* (Session D).

The timing of questionnaire-filling might matter for our test of reciprocity (Proposition 2), because hindsight bias could affect first-movers' stated expectations if they are asked to report what they had originally expected only once they have already observed second-movers' returns. This is why we add the *ex ante* design of Session D.[32] Finally, half the first movers of Session D were randomly assigned "bonus" status, meaning that they were told they would earn an additional 10 francs if the expected return $\hat{r}_j$ stated on their questionnaire turned out to be exactly correct.[33]

[31] 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 what we aim for, this design appears particularly attractive to our purpose, and we compare results of within- and between-subjects protocols. For an application of the within-subject design to control for confusion and warm-glow effects, see Palfrey and Prisbrey (1997).

[32] An additional implication of the *ex ante* design is that it could heighten first-movers' awareness of the second stage of the game, and thus their "rationality" (Croson, 2000).

[33] Camerer and Hogarth (1999), in a cross-experiment comparative study, report that incentives may improve subjects' performance in prediction tasks. Note that the bonus might induce first movers to make smaller transfers,

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 7.04 of their 10 francs to second movers, and second movers on average returned 11.02 francs. As a point of comparison, two trust games played with US students and both y_i and y_j of 10 dollars yielded averages sent (returned) of 5.16 (4.66) dollars (Berg *et al*., 1995) and 5.97 (4.94) dollars (Cox, 2003). Our sample subjects therefore appear to be extraordinarily trusting and trustworthy. Temporarily putting aside our caveats regarding cross-treatment comparability, we can interpret this as confirmatory evidence for Switzerland as a relatively "high-trust" society (see, e.g., Zak and Knack, 2001).

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 (11% of first movers and 20% of second movers across the four 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.

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 0.01 which has the "wrong" sign and is statistically insignificant. Virtually the same result obtains when we restrict the estimation to the within-subject protocols of Sessions B and D, controlling for first-mover 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.

---

as this reduces the number of potential amounts returned. This potential bias is controlled for via Session fixed effects in our regressions.

Next, we estimate a multi-group version of equation (6) by controlling for group-specific attributes that might affect mean transfers.[34] 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; etc.). Table 1 reports summary statistics on the distribution of those attributes in our subject sample. Estimation results are given in column III of Table 2. The coefficient on y_j is 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.[35] We find, however, that non-economics students send significantly more than economics and business majors.[36]

Finally, Sessions B and D have yielded significantly lower mean first-mover transfers than Session A (the omitted category). The low average transfers in Session D might be explained by the fact that the bonus payment to first movers induces them to transfer less in order to limit the number of possible amounts returned, and thus to heighten their chance of guessing the amount returned correctly. There is

[34] 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 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.

[35] 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.

[36] 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". The generalizability of this result outside the laboratory, however, is a moot point (see Yezer, Goldfarb and Poppen, 1996).

no ready explanation for the low transfers in Session B. This highlights the potential biases that would affect the between-treatment comparisons our study approach is designed to eschew.

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 estimates. We now find that the coefficient on y_j has the "correct" negative sign, but it continues to be statistically insignificant.[37] More importantly, we find none of the interaction effects to be statistically significant. It is particularly revealing that not even the interaction terms for Sessions B and D are significant, recalling that those were the sessions featuring the within-subject protocol and any impact of y_j on s_i could be expected to be particularly strong there.

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

In sum, our results thus far 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 Trust?

Table 4 reports regression estimates based on equation (8), in univariate form (column I) and with group-specific controls (column II). The estimated coefficient on $\hat{\rho}_j$ is positive, as expected, but statistically insignificant and economically small: the coefficient implies that a first mover who expects the second mover to return 1.5 times s_i, earning her a 50% "profit", sends only 0.05 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.

Some measurement error may be related to the mechanism of expectations revelation: *ex post* statements of *ex ante* expectations might be affected by hindsight bias, and lack of incentives could exacerbate the randomness in stated expectations. Figure 2 plots observed returns against first-movers' stated expectations and shows that stated expectations are in fact not correlated with observed returns. Regression (II) in Table 4 includes dummy variables for the timing of expectations revelation (*Exante_estimate* = 1 for Sessions A, B and C) and for the availability of a bonus for correct guesses (*Bonus* = 1 for the "bonus" group in Session D). Neither the main effects of these variables nor their interactions with $\hat{\rho}_j$ yield statistically significant coefficient estimates. Changes in the experimental protocol designed to improve the accuracy of stated first-mover expectations therefore have no discernible effect on the results obtained.

Another approach to limit the distorting impact of mismeasurement is to drop observations for which inaccurate reporting appears particularly probable. Hence, 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 five 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

[37] *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.

significant determinant of first-mover transfers. We also find that the quantitative impact of expected reciprocity increases more than eight-fold when we drop the five highly implausible observations, from 0.10 (column II) to 0.91 (column IV). The estimated coefficient of the full specification suggests that a first mover who expects the second mover to return 1.5 times s_i, earning her a 50% "profit", sends 0.46 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.[38] Finally, we restrict the sample to Sessions B and D and estimate within-subject parameters by including first-mover-specific fixed effects (Table 4, column V). This specification yields a statistically significant estimated coefficient on $\hat{\rho}_j$ of 1.57 - the trust hypothesis is once again supported.

One feature of all our regression results is the small share of observed variance in first-mover transfers that our regression models manage to explain. The highest R-squared is 0.22 (Table 4, column V).[39] This means that more than three quarters of the variation in "trust-like" transfers remains unexplained - even within subjects. This might be interpreted as a harsh indictment of our behavioral model. We instead regard it as a reminder that randomness looms large even in carefully controlled experimental settings, and that empirical tests of theoretical priors should pay explicit attention to the implications of such randomness (particularly to the danger of biased estimation).

---

[38] We also estimated the model with the full set of interactions (analogously to equation (7)). None of the interaction terms was found to be statistically significant. In addition, 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 using the Tobit estimator, but we again 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.

[39] This is the "within" R-square, expressing how much of the variance in first-mover transfers is explained by the model, once subject-specific effects are controlled for. The total R-square, considering also the explanatory power of the subject-specific effects, is 0.94.

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.

5.3 Or Is It Trust *and* Altruism?

So far, we have assumed that Propositions 1 and 2 are non-overlapping can be tested perfectly independently. Inspection of conditions (3a) and (3b) that underlie Proposition 1 shows that the two tests are not in fact fully separable. These conditions only cover the extreme cases of *only* altruism and *no* altruism. Our results reported above reject the "only altruism" hypothesis. However, a comprehensive test needs to allow for the simultaneous presence of altruism and other behavioral motives. First movers could be *partly* motivated by altruism and still send more to richer second movers, if their altruism is strong enough only to reduce but not fully to offset the gap between what they send to the rich and to the poor second movers. To test this more general version of Proposition 1, we need to establish whether there is a negative relationship between second-mover wealth and first-mover transfers *while controlling for the effect of trust motives*.

This is done in our regressions reported in Table 4, where we estimate coefficients on y_j, holding $\hat{\rho}_j$ fixed. Interestingly, we now observe the "correct" negative sign on y_j in four out of the five regressions, and the "within" specification (V) yields a (borderline) statistically significant coefficient estimate on y_j. There is thus some evidence in favor of the altruism hypothesis, once we control for trust. If we compare the two coefficients, however, we find that the trust motive dominates strongly. The standardized ("beta") regression coefficient on $\hat{\rho}_j$ (0.44) is more than twice as large in absolute value than the standardized coefficient on y_j (-0.21). Recall in addition that our estimated coefficient on $\hat{\rho}_j$ likely suffers from attenuation bias, and note that the RESET test suggests this regression to be misspecified.

We interpret this result as suggesting the existence of a subsidiary role for altruistic preferences, strongly dominated by reciprocity-based motives.

6. CONCLUSIONS

We propose discriminatory criteria to identify altruism and trust as determinants of first-mover transfers in trust games. The tests, formally derived from a model of first-mover preferences, are based on within-treatment and, in some experimental sessions, within-subject comparisons. They should therefore be immune to the experimental bias problem associated with the random component in the choices of laboratory subjects. Anonymity-preserving 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, accept expected reciprocation as dominant explanations for "trust-like" transfers. Altruism is rejected as a dominant explanation, but weakly supported as a subsidiary motivation.

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 (Ben-Ner *et al*., 2004; Clark and Sefton, 2000; Cox, Friedman and Gjerstad 2007; 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 (where trust is defined as a mixture of selfish payoff-maximization and intrinsic reward from co-operating with a fellow subject who is expected to be trustworthy). 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).

Throughout this study, we have insisted on the simple point that randomness 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 and unobservable subject-specific heterogeneity. This does not undermine the validity of economic experiments, but it calls for carefully designed hypothesis tests.[40]

This research could be extended in a number of ways. 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. Our observed first-mover transfers and second-mover returns are considerably higher than the average trust-game transfers reported in the literature. This conforms with prior behavioral findings for Switzerland, and it need not imply that our subjects are unrepresentative with respect to what matters to us, i.e. the *relative* utility weight of altruism and expected reciprocation. Nonetheless, it might be interesting to explore the robustness of our results by replicating the experiment with different subject pools. 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 study is firmly rooted in the traditional epistemological method of economics: formulate a rigorous but specific model of preferences, derive refutable hypotheses, and test these hypotheses on data collected from observed choices. The progress of modern neuroscience allows researchers to study the biological roots of human preferences. The finding by Kosfeld *et al*. (2005) that trust-game first-movers send more after they have inhaled a hormone known to enhance social attachment in animals provides an example of the opportunities offered by this approach. However, as long as preference components such as trust and altruism cannot be traced to different brain regions or

[40] See also Manski (2002) on the problem of identification and inference in experimental research.

individually activated by particular substances, we shall have to rely on tests that are based on preferences as revealed through observable economic choices.[41]

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.

Andreoni, James (1990) "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving". *Economic Journal*, 100(401): 464-477.

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.

Ashraf, Nava; Bohnet, Iris and Piankov, Nikita (2006) "Decomposing Trust and Trustworthiness". *Experimental Economics*, 9: 193-208.

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.

Bohnet, Iris and Zeckhauser, Richard (2004) "Trust, Risk and Betrayal". *Journal of Economic Behavior and Organization*, 55: 467-484.

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. (2002) "Swift Neighbors and Persistent Strangers: A Cross-Cultural Investigation of Trust and Reciprocity in Social Exchange". *American Journal of Sociology*, 108(1): 168-206.

[41] Note, however, that neuroscience seems to be making inroads here too, for example by providing indications that the brain's insula cortex is the neural home of fairness preferences (Camerer, Loewenstein and Prelec, 2005).

Buchan, Nancy R.; Johnson, Eric J. and Croson, Rachel T.A. (2006) "Let's Get Personal: An International Examination of the Influence of Communication, Culture and Social Distance on Other Regarding Preferences". *Journal of Economic Behavior and Organization*, 60: 373-398.

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.

Camerer, Colin F; Loewenstein, George and Prelec, Drazen (2005) "Neuroeconomics: How Neuroscience Can Inform Economics". *Journal of Economic Literature*, 43(1): 9-64.

Carter, Michael R. and Castillo, Marco (2006) "Trustworthiness and Social Capital in South Africa: An Experimental Approach". *Mimeo*, University of Wisconsin-Madison.

Charness, Gary (2004) "Attribution and Reciprocity in an Experimental Labor Market". *Journal of Labor Economics*, 22: 665-688.

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.

Charness, Gary and Rabin, Matthew (2002) "Understanding Social Preferences with Simple Tests". *Quarterly Journal of Economics*, 117(3): 817-869.

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., Friedman, Daniel and Gjerstad, Steven (2007) "A Tractable Model of Reciprocity and Fairness". *Games and Economic Behavior*, 59: 17-45.

Cox, James C.; Sadiraj, Klarita and Sadiraj, Vjollca (2007) "Implications of Trust, Fear, and Reciprocity for Modeling Economic Behavior". *Experimental Economics*, forthcoming.

Croson, Rachel T.A. (2000) "Thinking Like a Game Theorist: Factors Affecting the Frequency of Equilibrium Play". *Journal of Economic Behavior and Organization*, 41: 299-314.

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

Durlauf, Stephen N. (2002) "On the Empirics of Social Capital". *Economic Journal*, 112: F459-F479.

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

Eckel, Catherine C. and Wilson, Rick K. (2004) "Is Trust a Risky Decision?" *Journal of Economic Behavior and Organization*, 55: 447-465.

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; Klein, Alexander and Klaus M. Schmidt (2007) “Fairness and Contract Design”. *Econometrica*, 75(1): 121-154.

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.

Gächter, Simon; Herrmann, Benedikt and Thöni, Christian (2004) “Trust, Voluntary Cooperation, and Socio-Economic Background: Survey and Experimental Evidence”. *Journal of Economic Behavior and Organization*, 55: 505-531.

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

Greig, Fiona and Bohnet, Iris (2005) “Is There Reciprocity in a Reciprocal-Exchange Economy? Evidence from a Slum in Nairobi, Kenya”. *Mimeo*, Harvard University.

Guiso, Luigi; Sapienza, Paola and Zingales, Luigi (2007) “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.

Holm, Håkan J. and Danielson, Anders (2005) “Tropic Trust Versus Nordic Trust: Experimental Evidence from Tanzania and Sweden”. *Economic Journal*, 115: 505-532.

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.

Kosfeld, Michael; Heinrichs, Markus; Zak, Paul J.; Fischbacher, Urs and Fehr, Ernst (2005) “Oxytocin Increases Trust in Humans”. *Nature*, 435(2): 673-676.

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

Levitt, Steven D. and List, John A. (2007) “What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?” *Journal of Economic Perspectives*, 21(2): 153-174.

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.

Manski, Charles F. (2002) “Identification of Decision Rules in Experiments on Simple Games of Proposal and Response”. *European Economic Review*, 46: 880-891.

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.

Palfrey, Thomas R. and Prisbrey, Jeffrey E. (1997) “Anomalous Behavior in Public Goods Experiments: How Much and Why?” *American Economic Review*, 87(5): 829-846.

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

Rabin, Matthew (2000) “Risk Aversion and Expected-Utility Theory: A Calibration Theorem”. *Econometrica*, 68(5): 1281-1292.

Samuelson, Larry (2005) “Economic Theory and Experimental Economics”. *Journal of Economic Literature*, 43(1): 65-107.

Sapienza, Paola; Toldra, Anna and Zingales, Luigi (2007) “Understanding Trust”. *NBER Working Paper* #13387.

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

Sobel, Joel (2005) “Interdependent Preferences and Reciprocity”. *Journal of Economic Literature*, 43: 392-436.

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.

Yezer, Anthony M.; Goldfarb, Robert S. and Poppen, Paul J. (1996) “Does Studying Economics Discourage Cooperation? Watch What We Do, Not What We Say or How We Play”. *Journal of Economic Perspectives*, 10(1): 177-186.

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

Table 1: Data Description

	Session A	**Session B**	**Session C**	**Session D**	**TOTAL**
No. of observations	38	36[#]	31	64[#]	169
s_i [##]	7.76 (2.63)	6.44 (3.49)	6.77 (4.31)	7.08 (3.35)	7.04 (3.44)
occurrences of $s_i = 0$	0 (0%)	4 (11%)	7 (23%)	7 (11%)	18 (11%)
r_j [##]	12.37 (10.93)	8.06 (8.19)	10.45 (13.93)	12.17 (10.25)	11.02 (10.82)
occurrences of $r_j = 0$	4 (11%)	7 (19%)	12 (39%)	10 (16%)	33 (20%)
$\hat{\rho}_j = \hat{r}_j / s_i$ [##]	2.00 (1.12)	1.86 (0.95)	1.32 (0.82)	1.49 (0.71)	1.67 (0.92)
y_j [###]	13 * 0 12 * 10 13 *20	18 * 0 18 * 20	16 * 0 15 * 20	32 * 0 32 * 20	79 * 0 12 * 10 78 * 20
y_i [##]	10 (0)	10 (0)	10 (0)	10 (0)	10 (0)
Female	18.4%	38.9%	19.4%	40.6%	31.4%
Nat_Swiss	92.1%	83.3%	83.9%	87.5%	87.0%
Lang_French	81.6%	77.8%	77.4%	71.9%	76.3%
Lang_German	13.2%	11.1%	9.7%	12.5%	11.8%
Non-economist	2.6%	11.1%	0.0%	65.6%	27.8%

[#] In Session B(D), 36(64) observations correspond to 18(32) players 1, each matched with two players 2.
[##] Mean values (standard deviations in parentheses)
[###] (number of observations × *y*)

Table 2: Altruism Regressions, OLS
(dependent variable = s_i) #

	(I)	(II) ##	(III)	(IV)
y_j	0.01 (0.03)	0.003 (0.02)	0.01 (0.03)	-0.02 (0.11)
Female			-0.16 (0.58)	0.32 (0.86)
Female × y_j				-0.04 (0.06)
Nat_Swiss			-0.05 (0.91)	-0.41 (1.25)
Nat_Swiss × y_j				0.04 (0.10)
Lang_French			0.39 (0.81)	0.71 (1.18)
Lang_French × y_j				-0.03 (0.08)
Lang_German			0.83 (1.04)	1.39 (1.38)
Lang_German × y_j				-0.05 (0.10)
Non-economist			1.61 (0.75)**	1.37 (1.10)
Non-economist × y_j				0.02 (0.08)
Session_B			-1.39 (0.74)*	-1.94 (1.16)*
Session_B × y_j				0.05 (0.09)
Session_C			-0.90 (0.90)	-1.72 (1.33)
Session_C × y_j				0.08 (0.10)
Session_D			-1.62 (0.81)**	-1.81 (1.22)
Session_D× y_j				0.01 (0.09)
Dummies for 1st-movers	No	Yes	No	No
R-squared	0.001	0.0003	0.046	0.057
F statistic	0.12	0.02	1.08	0.67
Breusch-Pagan test ###	0.93	n.a.	0.10	0.07
RESET test ####	0.24	0.90	0.62	0.48
Observations	169	100	169	169

heteroskedasticity-consistent standard errors in brackets: *: 90% confidence level, **: 95% , ***: 99%; constant term included but not reported

fixed-effects panel data model; regression includes only observations from Sessions B and D; R-squared and F statistic only on "within" variation

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
(dependent variable = s_i) #

	(I)	(II) ##	(III)	(IV)
y_j	0.02 (0.06)	-0.003 (0.04)	0.02 (0.06)	-0.09 (0.24)
Female			-0.17 (1.32)	0.89 (1.84)
Female × y_j				-0.09 (0.14)
Nat_Swiss			-0.34 (2.03)	-1.07 (2.92)
Nat_Swiss × y_j				0.07 (0.21)
Lang_French			1.54 (1.72)	1.61 (2.55)
Lang_French × y_j				-0.003 (0.17)
Lang_German			2.24 (2.29)	2.52 (3.19)
Lang_German × y_j				-0.04 (0.22)
Non-economist			3.50 (1.74)**	2.40 (2.48)
Non-economist × y_j				0.11 (0.17)
Session_B			-2.77 (1.68)	-3.86 (2.61)
Session_B × y_j				0.10 (0.19)
Session_C			-1.34 (2.14)	-3.77 (3.00)
Session_C × y_j				0.25 (0.24)
Session_D			-3.62 (1.88)*	-3.68 (2.79)
Session_D × y_j				-0.005 (0.20)
Dummies for 1st-movers	No	Yes	No	No
Constant	8.99 (0.92)***	7.81 (0.69)***	9.13 (2.31)***	10.41 (3.21)***
Pseudo R-squared ###	0.0002	n.a.	0.0114	0.0156
Observations	169	100	169	169

two-sided Tobit with censoring at s_i =0 and s_i = 10; heteroskedasticity-consistent standard errors in brackets; *: 90% confidence level, **: 95%, ***: 99%

fixed-effects Tobit model; regression includes only observations from Sessions B and D.

= 1 - L_1/L_0, where L_0 and L_1 are the constant-only and full model log-likelihoods respectively

Table 4: Reciprocity Regression, OLS
(dependent variable = s_i) #

	(I)	(II)	(III)	(IV)	(V) ##
Observations included if:	$s_i > 0$		$s_i > 0,\ \hat{\rho}_j \le 3$		
$\hat{\rho}_j$ ###	0.09 (0.27)	0.10 (0.34)	0.67 (0.37)**	0.91 (0.40)**	1.57 (0.70)**
y_j	0.002 (0.02)	-0.004 (0.02)	-0.003 (0.02)	-0.005 (0.02)	-0.05 (0.03)*
Exante_estimate		-0.44 (1.35)		0.71 (1.41)	
Exante_estimate × $\hat{\rho}_j$		0.30 (0.61)		-0.30 (0.73)	
Bonus		-0.51 (1.83)		-0.20 (1.88)	
Bonus × $\hat{\rho}_j$		0.06 (1.11)		-0.17 (1.15)	
Female		-0.91 (0.51)*		-0.98 (0.50)*	
Nat_Swiss		-0.43 (0.67)		-0.72 (0.66)	
Lang_French		1.12 (0.86)		1.19 (0.75)	
Lang_German		1.40 (0.86)		1.46 (0.88)*	
Non-economist		0.89 (0.56)*		0.98 (0.55)*	
Session_B		-0.54 (0.67)		-0.19 (0.62)	
Session_C		0.98 (0.68)		1.30 (0.65)**	
Dummies for 1st-movers	No	No	No	No	Yes
Constant	7.76 (0.53)***	7.21 (1.02)***	6.98 (0.58)***	6.02 (1.02)***	5.87 (0.89)***
Observations ◊	150	150	145	145	86
R-squared	0.001	0.112	0.033	0.152	0.220
F statistic (full model)	0.06	1.60*	1.97	1.99**	2.54*
Breusch-Pagan test ◊◊	0.80	0.05	0.05	0.01	n.a.
RESET test ◊◊◊	0.02	0.22	0.31	0.04	0.00

heteroskedasticity-consistent standard errors in brackets; *: 90% confidence level, **: 95%, ***: 99%

fixed-effects panel data model; regression includes only observations from Sessions B and D; R-squared an *F* statistic only on "within" variation.

confidence levels with respect to one-tailed *t* test

◊ 18 observations with $s_i = 0$ and one observation with unreported $\hat{r}_j$ had to be 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

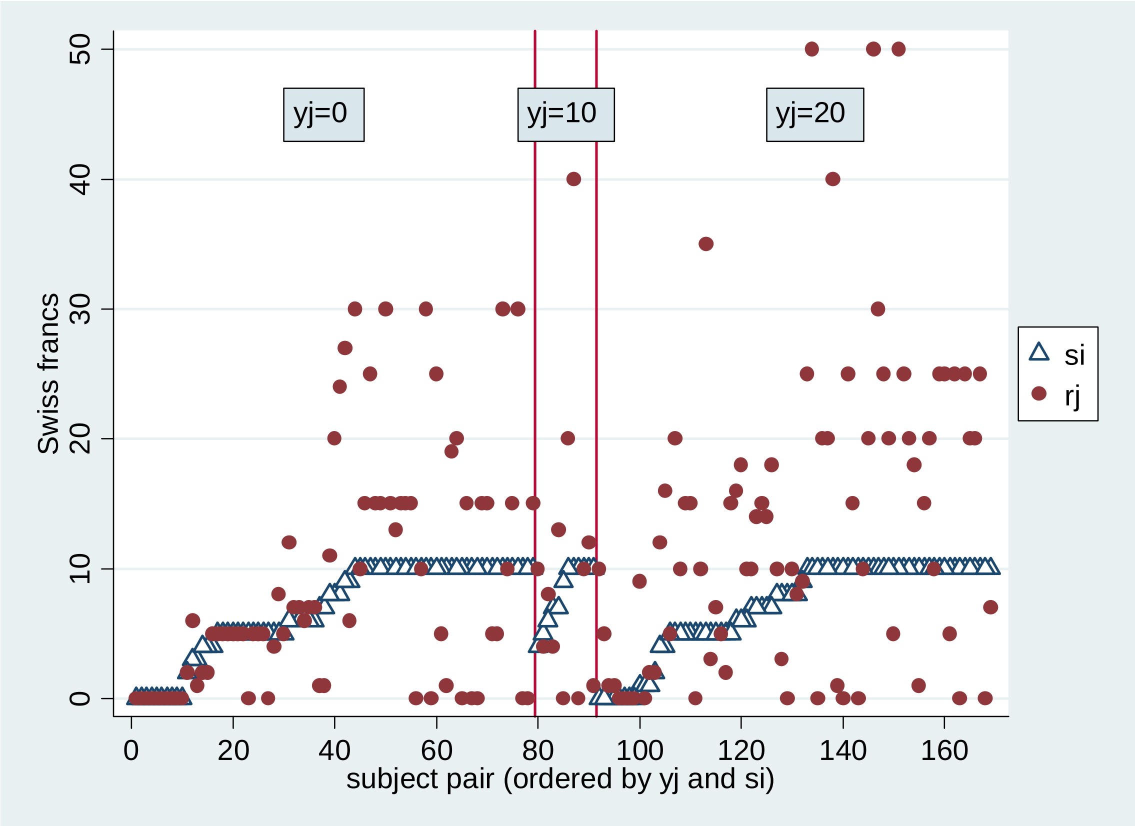
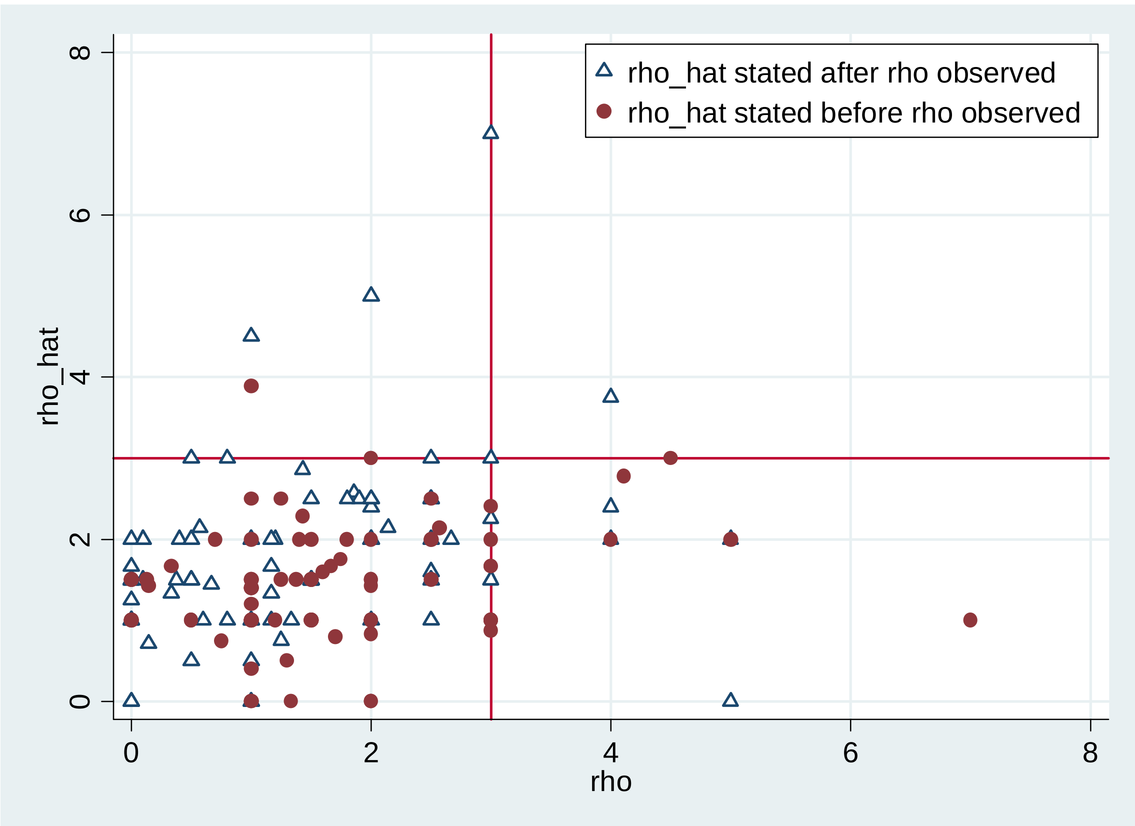


Figure 2: Expected and Actual Second-Mover Returns



APPENDIX: IS EXPECTED RECIPROCATION BALANCED?

According to our assumption (1f), first movers expect second-movers' rate of return ($\hat{\rho}_j$) to be independent of the amount sent (s_i). Greig and Bohnet (2005) term this assumption "balanced" expected reciprocation (BER), while "conditional" expected reciprocation (CER) implies that rates of return vary with amounts sent. BER is important to our analysis primarily because it allows consistent estimation of our reciprocity regressions. CER would inevitably lead to simultaneity bias.

A simple test for BER consists of regressing first movers' expected amount returned ($\hat{r}_j$) on the amount sent (s_i) and the square of the amount sent (s_i^2). If we abstract from potential non-linearity in risk aversion with respect to expected returns, BER implies that the estimated coefficient on the square term is zero.[42] A coefficient on the square term different from zero would imply CER.

We run this test for our sample of first movers that reported plausible expected rates of returns ($\hat{\rho}_j \leq 3$)[43] in three variants: a parsimonious model that includes only s_i and s_i^2, a model that adds the available group-level controls, and a within-subject specification. The results are reported in Table A1. All three regressions return statistically insignificant coefficients on s_i^2, thus supporting BER.[44]

[42] If s_i were exogenous, BER would furthermore imply that the estimated coefficient on the amount sent is equal to one. Since reciprocity implies causation running from expected returns to amounts sent (Proposition 2), and since our data contain actual amounts sent (as opposed to hypothetical amounts suggested to first movers), we must expect estimated regression coefficients to exceed unity in our empirical setting.

[43] Our results carry through fully if we include all observations.

[44] Running the same regression on data based on trust games played in Kenya, Greig and Bohnet (2005) also support BER and reject CER.

Appendix Table 1: Testing for Balanced Expected Reciprocation
(OLS, dependent variable = $\hat{r}_j$)[#]

	(I)	(II)	(III)[##]
s_i	2.10 (0.62)***	1.52 (0.70)**	0.30 (1.08)
s_i^2	-0.02 (0.05)	0.02 (0.05)	0.11 (0.10)
y_j		0.19 (0.04)***	
Exante_estimate		-0.01 (0.57)	
Bonus		-2.03 (1.11)*	
Female		-1.04 (0.91)	
Nat_Swiss		1.32 (1.21)	
Lang_French		0.23 (1.35)	
Lang_German		-1.48 (1.52)	
Non-economist		-2.10 (0.90)**	
Session_B		0.40 (1.26)	
Session_C		-4.04 (1.73)**	
Dummies for 1st-movers	No	No	Yes
Constant	-2.39 (1.63)	15.84 (1139.3)	2.45 (3.87)
Observations [###]	145	145	86
R-squared	0.39	0.53	0.23
F statistic (full model)	112.8	25.80***	5.21**
Breusch-Pagan test ◊	0.00	0.00	n.a.
RESET test ◊◊	0.95	0.27	0.37

[#] heteroskedasticity-consistent standard errors in brackets; *: 90% confidence level, **: 95% confidence level, ***: 99% confidence level
[##] fixed-effects panel data model; regression includes only observations from Sessions B and D; R-squared and *F* statistic only on "within" variation
[###] estimation restricted to observations with $\hat{\rho}_j \leq 3$
◊ *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)