Does the trust game measure trust?

Marius Brülhart [a,b,*], Jean-Claude Usunier [a]

[a] *University of Lausanne, Switzerland*
[b] *CEPR, Switzerland*

ARTICLE INFO

ABSTRACT

Could altruism explain observed choices in the standard trust game? With dominant altruism, trustors would give more to poor trustees. However, we find trustors to give no more to poor than to rich trustees, confirming trust as the dominant motivation for "trust like" choices.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The "trust game" of Berg et al. (1995) has become the standard laboratory experiment for measuring trust. In the trust game, a *first mover* is randomly and anonymously paired with a *second mover*. Both are given a monetary endowment. The first mover may transfer some or all of his endowment to the second mover. This transfer is tripled by the experimenter and handed to the second mover, and finally the second mover may return some or all of the received transfer. First-mover transfers are interpreted as a manifestation of trust, and second-mover transfers as a manifestation of trustworthiness.

We propose an amended protocol as a way to investigate the claim that *altruism* might play a significant role in "trust like" first-mover transfers (Smith, 2003; Cox, 2004). 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 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.

A further potential complication is that there may be idiosyncrasies in individual preferences and, more importantly, 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 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.[1]

2. Experimental details

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.[2] 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 amount r_j, $0 \leq r_j \leq (y_j + 3s_i)$, to their

---

[1] For a between-subject and between-treatment test of trust versus altruism, see Cox (2004).

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

Table 1
Data description.

	Session A	Session B	Session C	Session D	Total
No. of observations	38	36[a]	31	64[a]	169
s_i[b]	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[b]	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%)
	13 * 0				79 * 0
y_j[c]	12 * 10	18 * 0	16 * 0	32 * 0	12 * 10
	13 * 20	18 * 20	15 * 20	32 * 20	78 * 20
y_i[b]	10 (0)	10 (0)	10 (0)	10 (0)	10 (0)

[a] In Session B (D), 36 (64) observations correspond to 18 (32) players 1, each matched with two players 2.
[b] Mean values (standard deviations in parentheses).
[c] (Number of observations *y).
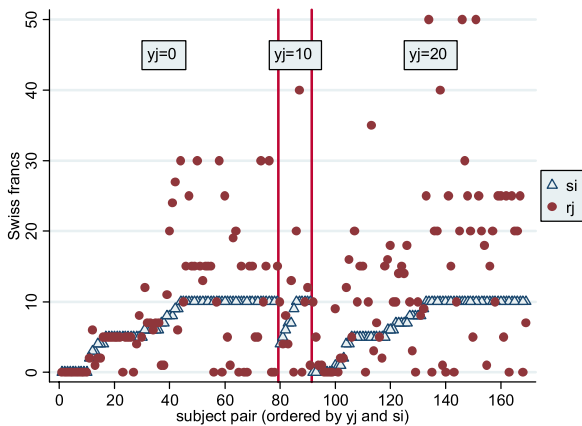
Fig. 1. First-mover transfers and second-mover returns.

paired first movers.[3] 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 $Y_j = y_j + 3s_i - r_j = y_j + s_i(3 - \rho_j)$. We write first movers' expected rate of return from second movers as $\hat{\rho}_j$.

In the original interpretation of trust-game choices, first-mover transfers are a manifestation of trust. Trust, in turn, has two components. One component is due to "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). The other component is due to selfishness and the expectation of positive reciprocation. Hence, 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.

In a longer version of this paper, we develop an explicit model of subject motivations in such an experimental setting (Brülhart and Usunier, 2008).[4] This allows us to derive testable hypotheses formally from a precisely specified utility function. Given their intuitive plausibility, we concentrate on verbal statements of our central propositions here.

We vary second movers' initial wealth y_j and then examine the relationship between s_i and y_j. Given the double-anonymous experimental protocol, y_j constitutes the only element of information that first movers have about their paired second movers. First movers will 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 are able to "afford" a higher ρ_j. If, except for randomness, altruism is first-movers' sole motivation, they will give more to poor second movers than to rich ones. Conversely, first-mover transfers in the absence of own altruistic motives will be non-negatively related to second-mover wealth — in line with the original interpretation of trust-game results.

We can therefore formulate the following discriminating hypothesis, based entirely on observable variables.

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 either send less to poor second movers than to rich second movers, or they send equal amounts to both.*

We have played the trust game with first-year 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.[5] 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-anonymous procedures.[6] The four sessions were organized such as to offer variation in two methodological dimensions: manual game (Sessions A and B) versus internet-based game (Sessions C and D); and between-subject design (Sessions A and C) versus within-subject design (Sessions B and D).

3. Results

Summary statistics of observed transfers are reported in Table 1 and illustrated in Fig. 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. Furthermore, our experiments confirm the finding that only a small fraction of

---

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

[4] The main assumptions of that model are (i) that selfish, altruistic, reciprocal and random motivations enter agents' utility in additively separable fashion, (ii) that utility over own income is concave with nonincreasing absolute risk aversion, (iii) that altruism incorporates an element of inequality aversion, and (iv) that first movers' expected rate of return and the variance thereof are independent of the amount sent ("balanced reciprocation" in the terminology of Greig and Bohnet (2008)).

[5] One Swiss franc was worth approximately 0.73 and 0.85 US dollars at the time of the experiments.

[6] The texts of the "recruitment email", experimental instruction sheets and the post-experiment questionnaire can be obtained from the authors on request. Further details are given in Brülhart and Usunier (2008).

Table 2
Regression results (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
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* to *D*; *R*-squared and *F* statistic only on "within" variation.

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.

We test Proposition 1 by regressing s_i on y_j using ordinary least squares.[7] 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.

Next, we estimate a multi-group version of Proposition 1 by controlling for group-specific attributes that might affect mean transfers. We consider five attributes: gender, nationality, native tongue, subject of study, and experimental session. 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. We find, however, that non-economics students send significantly more than economics and business majors.

Finally, Sessions *B* and *D* have yielded significantly lower mean first-mover transfers than Session *A* (the omitted category). There is no ready explanation for these differences. 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. 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.[8] 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.

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

4. Conclusions

We propose a discriminatory criterion to identify altruism and trust as determinants of first-mover transfers in trust games. The test is based on within-treatment and within-subject comparisons and should therefore be immune to the experimental bias problem associated with the random component in the choices of laboratory subjects. Using data gathered in experiments 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 a dominant determinant of "trust-like" decisions.

A 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.[9]

---

[7] To account for the two-sided censoring of s_i implied by the trust game, we have also estimated the four equations using the two-sided Tobit estimator. The results are qualitatively unchanged from the OLS runs.

[8] *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.

[9] Johansson-Stenman et al. (2005) find that stake size may affect not just the dispersion but also the mean of trust-game transfers.

References

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

Brülhart, M., Usunier, J.-C., 2008. Verified trust: reciprocity, altruism, and randomness in trust games, IRM Working Paper #0809, Institute of Research in Management, University of Lausanne.

Cox, J.C., 2004. How to identify trust and reciprocity. Games and Economic Behavior 46, 260–281.

Greig, F., Bohnet, I., 2008. Is there reciprocity in a reciprocal-exchange economy? Evidence from a slum in Nairobi, Kenya. Economic Inquiry 46, 77–83.

Johansson-Stenman, O., Mahmud, M., Martinsson, P., 2005. Does stake size matter in trust games? Economics Letters 88, 365–369.

Smith, V.L., 2003. Constructivist and ecological rationality in economics. American Economic Review 93, 465–508.

Sobel, Joel, 2005. Interdependent preferences and reciprocity. Journal of Economic Literature 43, 392–436.