Rationality and Society
http://rss.sagepub.com/

Institutions and Trust: Implications for Preferences, Beliefs and Behavior

Iris Bohnet and Yael Baytelman

Rationality and Society 2007 19: 99
DOI: 10.1177/1043463107075110

The online version of this article can be found at:
http://rss.sagepub.com/content/19/1/99

Published by:
SAGE
http://www.sagepublications.com

INSTITUTIONS AND TRUST

IMPLICATIONS FOR PREFERENCES, BELIEFS AND BEHAVIOR

Iris Bohnet and Yael Baytelman

ABSTRACT

Institutions matter – but how? This article employs experiments to examine whether institutions only affect trust and trustworthiness behavior by changing constraints and thereby the beliefs people hold about others' behavior, as commonly assumed in a rational choice framework, or also by influencing preferences. In a within-subject design, we confront people with an anonymous one-shot trust game, a one-shot game with pre-play communication, post-play communication and a post-play punishment option and a finitely repeated game. Institutions increasing the cost of betrayal as compared to an anonymous one-shot game affect people's beliefs and enhance their willingness to trust and be trustworthy. However, all settings that offer tighter institutional constraints compared to the anonymous one-shot game decrease intrinsically motivated trust. They do not influence trustworthiness. Thus, institutions may also affect preferences. The 'crowding-out' of intrinsic trust is of concern as it has been found to be associated with economic performance and democracy.

KEYWORDS • trust • trustworthiness • incentives • intrinsic motivation • crowding-out • experiments

1. Introduction

'For he that performeth first, has no assurance the other will perform after' (Hobbes 1960: 89). But those who are willing to take this risk and trust another person contribute to our well-being. Generalized trust is associated with economic growth (Knack and Keefer 1997; Zak and Knack 2001), stable democracy (Inglehart 1999), better functioning governments (LaPorta et al. 1997; Knack 2002), social capital (Putnam 1993, 2000), a decrease in crime (Rosenfeld et al. 2001) and cooperation within and between organizations (Kramer and Tyler 1996).

Rationality and Society Copyright © 2007 Sage Publications. Vol. 19(1): 99–135.
www.sagepublications.com DOI: 10.1177/1043463107075110

If good things happen when people trust each other, we may wonder how we can increase trust. We adopt a definition of trust recently proposed by a cross-disciplinary review as 'a psychological state composing the intention to accept vulnerability based on positive expectations of the intentions or behavior of another' (Rousseau et al. 1998: 395). This definition combines the notion of trust as a social motivation (Mansbridge 1999) with the view that trust is a belief (Hardin 2002). Thus, to enhance trust, we could try to affect people's willingness to be vulnerable and/or their expectations of trustworthiness.

This article examines whether institutions positively affecting people's beliefs about the returns to trust also influence their innate willingness to be vulnerable, and what effect this has on trust and trustworthiness behavior. We look at both, trust and trustworthiness. In all likelihood, institutions affecting beliefs about trustworthiness also influence actual trustworthiness (although beliefs may well not be completely accurate). We define trustworthiness as the willingness to reward trust at a cost to oneself, based on one's expectations of trust.

We introduce a new framework to measure trust and trustworthiness: both parties, the trustor and the trustee, simultaneously decide based on their expectations of their counterpart's behavior. This approach allows us to measure the effect of a given institution on people's expectations, and their social motivation to be vulnerable and accept (the risk of) monetary losses. A sequential game where the trustee responds to the trust offered by the trustor would not allow us to examine how institutions affect the trustee's beliefs and his response to these beliefs.[1]

In many real-life situations, trustors and trustees do not know their counterpart's move at the time they make a decision. Employers often decide whether to trust or to monitor their employee without knowing whether the employee will honor their trust or not. Employees often decide whether to be trustworthy or shirk without knowing how much trust they get from their employer – but they may well have formed expectations about their employer's likely level of trust. The same applies to spouses, medical doctors and patients, or attorneys and clients. It is relevant for the relationship between voters and politicians. Voters often decide about whether to trust a representative or not based on their expectations of his or her trustworthiness, not on his or her actual trustworthiness. And politicians often decide about their degree of trustworthiness without knowing how much trust the electorate places in them, based on their beliefs about the electorate's trust.

In a rational choice framework, institutions are expected to affect behavior by changing constraints but not by influencing motivation. In

contrast, using an evolutionary approach, Bohnet et al. (2001) present a theoretical model for how preferences might adapt to institutional environments. They show that institutions increasing the expected cost of betrayal and thus, expectations of trustworthiness, need not enhance trust and trustworthiness behavior. In addition to changing incentives, institutional constraints may 'crowd out' the intrinsic motivation to assume vulnerability and trust and be trustworthy.[2] Even if institutions increased observed trust and trustworthiness behavior on average but at the same time destroyed intrinsically motivated trust, they would erode an important component of what has been identified as the foundation of social interaction and of politics more specifically (Coleman 1990; Putnam 1993; Fukuyama 1995; Nye et al. 1997). Gambetta noted (1988: 219): 'If other people's actions were heavily constrained, the role of trust in governing our decisions would be proportionately smaller ...' and Sitkin and Roth (1993: 376) observed that 'legalistic remedies can erode the interpersonal foundations of a relationship they are intended to bolster because they replace reliance on an individual's goodwill with objective, formal requirements.'[3] Thus, institutions may not be 'motivation compatible.' An institution is said to be motivation compatible if it does not affect people's intrinsic motivation to do something, neither positively nor negatively.

Ultimately, the effect of institutions on motivation, beliefs and behavior is an empirical question. We use survey experiments to address it.[4] We confront 353 senior executives with five institutional environments: an anonymous one-shot trust game scenario without communication; a one-shot trust game scenario with face-to-face communication before making a decision; a one-shot trust game scenario with face-to-face communication after making a decision; a one-shot trust game scenario allowing for post-play monetary punishment of betrayal; and an anonymous finitely repeated trust game scenario. None of the institutions we study is new to the experimental literature but so far, their effect on intrinsic trust and trustworthiness has not been studied. To measure how a given person responds to different institutional environments, and what motivates a given person to do so, we use a within-subject design where the same person is exposed to different institutional environments.[5]

Our trust game scenario builds on the investment game (Berg et al. 1995) but we do not play the game sequentially. Trustors and trustees simultaneously decide based on their expectations of their counterpart's behavior. We measure the intrinsic motivation to assume vulnerability and trust and be trustworthy by a subject's 'social orientation.' Kramer (1999: 573) suggested: 'Trust needs to be conceptualized not only as a

calculative orientation toward risk, but also a social orientation toward other people and toward society as a whole,' and Mansbridge (1999: 294) noted the importance of altruistic trust: 'To qualify as altruistic trust my move must ... be motivated by a conscious or unconscious intent to benefit the other or an intent to uphold a principle that in the long run usually benefits others.'

Social orientation may be due to social preferences such as altruism (Andreoni and Miller 2002) and inequity aversion (Fehr and Schmidt 1999; Bolton and Ockenfels 2000) or due to internalized norms, i.e. the psychological benefits (costs) that an individual derives from being kind (unkind) to others (Andreoni 1990; Crawford and Ostrom 1995). To measure a person's social orientation, each subject also participates in a dictator game scenario (Kahneman et al. 1986). A small number of studies have examined the relevance of social orientation for trust and trustworthiness in one-shot games. Recent work by Dufwenberg and Gneezy (2000), Cox (2004) and Ashraf et al. (2006) suggests that for most people, trust and trustworthiness in one-shot games are also related to their social orientation.

In contrast to the earlier work on motivation and trust, this article analyzes the effect more complex environments than the one-shot game have on trust and trustworthiness and on the underlying motivations and beliefs. In her Presidential Address on a 'Behavioral Approach to the Rational Choice Theory of Collective Action,' Elinor Ostrom (1998: 12) noted: 'At the core of a behavioral explanation are the links between the trust that individuals have in others, the investment others make in trustworthy reputations, and the probability that participants will use reciprocity norms (p. 12) … An important set of questions is related to how institutions enhance or restrict the building of mutual trust, reciprocity, and reputation' (p. 17). This article aims to address some of these questions. It is organized as follows: section II introduces the experimental design and section III presents a conceptual framework. Section IV discusses the results and section V concludes.

2. Experimental Design

We employ a within-subject design. All subjects participated in five different trust game scenarios and completed a post-experimental questionnaire on basic demographic characteristics. In addition, trustors participated in a triple dictator game scenario and trustees in a dictator game scenario. We run survey experiments (Duch and Palmer 2004), in

which subjects are asked to indicate what they would do if they were confronted with a particular situation.[6]

The sequence in which subjects were confronted with the six decisions was varied. In half of our sessions, subjects made the (triple) dictator game decision first and then were exposed to the trust game scenarios; in the other half, they first participated in the trust games and then in one of the dictator game scenarios. In the trust game scenarios, people first participated in the anonymous one-shot treatment, then the pre-play communication treatment (five minutes of free discussion before making the decisions), then the post-play communication treatment (five minutes of free discussion after having made and learned about the decisions), the punishment treatment (post-play punishment option to decrease the trustee's payoffs after having made and learned about the decisions) and the repetition treatment (finitely repeated game for five rounds).

The main experiment was run with 353 senior executives attending executive programs at a major university in the USA.[7] Participation was voluntary. The program participants were invited by an experimenter (who had not taught in the class before) to take part in a research project in an additional session that lasted approximately 45 minutes. We conducted 10 different sessions in large classrooms with between 30 and 38 people present. We randomly assigned the trustor role to half and the trustee role to the other half of the subjects in a given session.[8]

To get a better sense for possible order effects, we had a control group of an additional 194 senior executives play one of the game scenarios only. Participating in all games sequentially may induce people to anchor on the first game they see, decreasing the differences in behavior between the various institutional settings. Alternatively, sequencing could lead to contrast effects as counterfactuals are available, increasing the differences in behavior between the various settings as compared to the control experiment. We had no prior on how the two experimental designs might compare.

In the dictator game scenario, each subject was asked to allocate a fixed endowment F between herself, the dictator, and a random anonymous recipient who had not received any money and who they did not know (i.e. not a person in their session). They were informed that the dictator would earn $F - \hat{S}$ and the recipient $\hat{S}$. The only difference between the standard dictator game and the triple dictator game scenarios was that the amount sent to the recipient was tripled by the experimenter in the latter. Accordingly, in the triple dictator game, the dictator's earnings were as before, $F - \hat{S}$ but the recipient's earnings

were $3\hat{S}$. We refer to the amounts passed along in the (triple) dictator game as *willingness to send,* $\hat{S}_{(T)DG}$.

The only difference between the triple dictator and the trust game scenarios was that the trustee could send back any amount of money $\hat{Y}$ between 0 and $3\hat{S}$. Subjects were told that the trustor's earnings would be $F - \hat{S} + \hat{Y}$ while the trustee would earn $3\hat{S} - \hat{Y}$. We refer to the amounts sent by the trustor as *willingness to trust*, $\hat{S}$, and the amounts returned by the trustee as *willingness to be trustworthy*, $\hat{Y}$. In addition to indicating how much they wanted to send, we also asked trustors to report how much they expected the trustee to return to them, $E(\hat{Y})$, and trustees what they expected the trustors to send to them, $E(\hat{S})$, in each game. We asked subjects about their beliefs at the same time as we asked them about their willingness to trust (be trustworthy).[9] We acknowledge that this is not a perfect measure of expectations as subjects basically decide about behavior and expectations jointly. We will address this as a problem of recursive equations below (Pindyck and Rubinfeld 1991: 298). If we find that institutions affect subjects' motivations and behavior even when controlling for a variable predicting expectations, this would provide particularly strong support for the crowding hypothesis.

F equaled to a hypothetical US$100 in all our games. Participants were not paid for performance. A non-standard experimental format was used to test our hypotheses – a within-subject design to test for crowding by controlling for subjects' social preferences and expectations; a non-standard subject pool to examine the role of demographic characteristics somewhat more globally than just among American student-subjects; and a hypothetical format without monetary stakes. The latter was a necessary compromise to be allowed to run experiments with executives. To get a sense for the effects of our design, we compare our results to more standard experimental formats run with student-subjects with economic incentives, employing a between-subjects design.

3. Conceptual Framework

As a baseline, we assume that both trustors and trustees are selfish money-maximizers, and that trustors are aware of their trustees' preferences. The unique Nash equilibrium in our anonymous one-shot game and the subgame perfect equilibrium in our finitely repeated game predict that people are not willing to trust or be trustworthy.[10] This prediction also applies to the pre-play and post-play communication games because 'talk is cheap' and selfish trustors and trustees are not expected

to care about the social norms that may be evoked during communication in a traditional rational choice framework (Farrell and Rabin 1996).

In contrast, the availability of costless post-game monetary sanctions changes the equilibrium prediction. Subjects were told that:

'This scenario is a one-shot occurrence and a third stage is added, such that after both #1 and #2 have decided how much money to send each other, #1 has the option to decrease the amount of money that #2 holds by 50%. If #1 chooses this option, #1 does not gain or lose any money; the money is simply taken away from #2.'

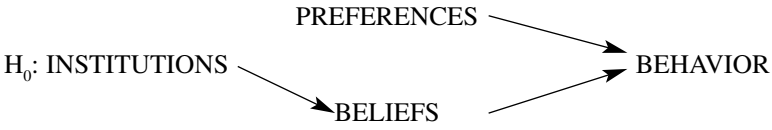
Selfish trustors are indifferent between punishing and not punishing (punishment does not affect their own payoffs); trustors with social preferences such as envy, inequity aversion or spite, punish. This leads to a multitude of possible equilibria. Fearing spiteful trustors, a selfish trustee who does not return any money if sent a positive amount in the anonymous one-shot game, may return a positive amount when threatened by monetary sanctions because his expected punishment costs outweigh the gains to be had if he kept everything.

In contrast to the baseline assumptions, trustors may send positive amounts in any of our games if they are motivated by intrinsic motivation – their social orientation – and/or if they expect social orientation to affect their trustee's decision. Trustees may return positive amounts if they are socially motivated. Substantial experimental evidence shows that even in the anonymous one-shot game, trust and trustworthiness levels far exceed the theoretically predicted levels. On average, trustors have been found to send about 50% and trustees to return about the amount sent in one-shot games (Camerer 2003). This strongly suggests that either some trustees are socially motivated and trustors anticipate this or that both some trustees and some trustors are socially motivated.

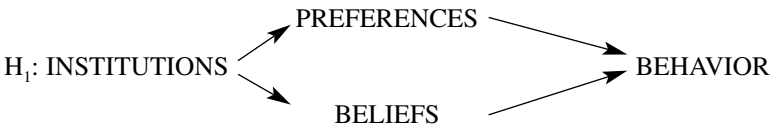
Communication and repetition add psychological and social costs and/or benefits to subjective payoffs. They induce similar dynamics as monetary sanctions, also leading to multiple equilibria (Granovetter 1992; Ostrom 1998). Such dynamics lead trustees to return larger amounts even when only constrained by social norms activated by informal institutional arrangements. Rational trustors anticipate this and respond by sending larger amounts. In between-subjects designs, typically comparing one of our institutional variations with the anonymous one-shot game, repetition and communication have generally been found to increase trust and cooperation behavior in trust games and related public goods or social dilemma experiments (for finitely repeated trust games, see e.g., Bohnet and Huck 2004, Engle-Warnick and Slonim 2004; for pre-play communication, e.g., Ostrom et al. 1992,

1994, and for a meta-analysis, Sally 1995; for post-play communication – studied less than our other institutional designs – Gächter et al. 1996 who only find a small effect).

We are interested in whether our institutions affect the relative importance of social motivation for trust and trustworthiness in our five games. Our *null hypothesis* is that institutions affect behavior only by changing beliefs about what others do, i.e. the amounts expected to be sent and returned, but not by influencing motivation:



In contrast, the *crowding hypothesis* suggests that in addition to affecting beliefs, institutional constraints also influence preferences, decreasing the relevance of intrinsic motivation for trust and trustworthiness:



For crowding to apply, we expect the amounts sent and returned in the trust game with institutional constraints to be less related to the amounts sent in the triple dictator game and the regular dictator game, respectively, than in the one-shot environment.

To test the two hypotheses, we first estimate the relationship between institutions (pre-communication, post-communication, punishment, repetition) and beliefs ($E(\hat{Y})$, the amount trustors expect to be returned, and $E(\hat{S})$, the amount trustees expect to be sent) in a first-stage regression. We include a number of demographic controls and also people's social orientation ($\hat{S}_{TDG,}$ the amount trustors send in the triple dictator game, and $\hat{S}_{DG}$, the amount trustees send in the dictator game):

$$E(\hat{Y})[E(\hat{S})] = \alpha + \beta*\text{precomm.} + \gamma*\text{postcomm.} + \delta*\text{punish.} + \varepsilon*\text{repet.} + \zeta*\hat{S}_{TDG[DG]} + \eta*\text{controls} \quad (1)$$

We then proceed by estimating the relationship between institutions and behavior ($\hat{S}$, the amount trustors are willing to send and $\hat{Y}$, the

amount trustees are willing to return), controlling for beliefs, social orientation and demographic variables. Interactions between the institutions and a person's social orientation tell us how relevant crowding is. The *crowding hypothesis* is supported if people are significantly less motivated to trust (reward trust) by their social orientation in the institutionally constrained treatments than in the anonymous one-shot game:

$$\begin{aligned}\hat{S}[\hat{Y}] = {} & \alpha + \beta^{*}\text{precomm.} + \gamma^{*}\text{postcomm.} + \delta^{*}\text{punish.} \\ & + \varepsilon^{*}\text{repet.} + \zeta^{*}\hat{S}_{TDG[DG]} + \eta^{*}\hat{S}_{TDG[DG]}{}^{*}\text{precomm.} \\ & + \theta^{*}\hat{S}_{TDG[DG]}{}^{*}\text{postcomm.} + \iota^{*}\hat{S}_{TDG[DG]}{}^{*}\text{punish.} \\ & + \kappa^{*}\hat{S}_{TDG[DG]}{}^{*}\text{repet.} + \lambda^{*}E(\hat{Y})[E(\hat{S})] + \mu^{*}\text{controls} \end{aligned} \quad (2)$$

If the institutions turn out to be strongly related to people's expectations in regression (1) but not to their willingness to trust and be trustworthy in regression (2), this is in line with the *null hypothesis*. Accordingly, we then use institutions as an instrument for beliefs, estimating in the second stage:

$$\hat{Y}[\hat{S}] = \alpha + \beta^{*}\hat{S}_{TDG[DG]} + \gamma^{*}\, E(\hat{Y})[E(\hat{S})] + \delta^{*}\text{controls (IV)} \quad (3)$$

If institutions do not only work through changing expectations but also affect behavior through other means, e.g. by changing preferences, we cannot use institutions as an instrument for expectations. At the same time, we still have a recursive equation system where endogenous variables can be determined sequentially (see Pindyck and Rubinfeld 1991: 298). In our case, beliefs can be determined first and then be used to predict behavior. We thus include the *predicted value for expectations*, based on the first-stage regression results, into the second-stage regression. This controls for the effect institutions, social orientation and demographic characteristics have on expectations and allows us to parse out the additional effect these variables might have on behavior:

$$\begin{aligned}\hat{S}[\hat{Y}] = {} & \alpha + \beta^{*}\text{precomm.} + \gamma^{*}\text{postcomm.} + \delta^{*}\text{punish.} \\ & + \varepsilon^{*}\text{repet.} + \zeta^{*}\hat{S}_{TDG[DG]} + \eta^{*}\hat{S}_{TDG[DG]}{}^{*}\text{precomm.} \\ & + \theta^{*}\hat{S}_{TDG[DG]}{}^{*}\text{postcomm.} + \iota^{*}\hat{S}_{TDG[DG]}{}^{*}\text{punish.} \\ & + \kappa^{*}\hat{S}_{TDG[DG]}{}^{*}\text{repet.} + \gamma^{*}\text{predicted } E(\hat{Y})[E(\hat{S})] \\ & + \mu^{*}\text{controls} \end{aligned} \quad (4)$$

The relative importance of crowding in our four treatments institutionally constraining behavior is an empirical question. Based on previous experimental evidence on the negative effects of sanctions (Ostrom et al.

1992; Bohnet et al. 2001; Fehr and Gächter 2002), we expect punishment to have more negative effects on intrinsic motivation than the more informal mechanisms activated through repetition and communication. Fehr and Gächter (2002) and Fehr and List (2004), for example, found that a punishment option is particularly effective for increasing trustworthiness when *not* used.

4. Results

We first discuss our data in the anonymous one-shot game scenario and compare them with standard experimental results in standard investment games (for monetary incentives) to get a sense for how the special features of our design might have affected results. We then turn to the various game scenarios with institutional constraints and more precisely analyze how institutions affect beliefs, preferences and willingness to trust and be trustworthy by running regressions. Table B1 in the Appendix contains the definitions of our main variables.

In the anonymous one-shot game scenario, trustors are willing to entrust their trustees with $\hat{S}$ = US$55 ($n$ = 158) and trustees to return $\hat{Y}$ = US$61 ($n$ = 182) to their trustors, on average. Trustors expect trustees to return $E(\hat{Y})$ = US$70 and trustees expect trustors to send $E(\hat{S})$ = US$42 on average, suggesting that trustors are slightly too optimistic and trustees slightly too pessimistic. In the triple dictator game, subjects are willing to send $\hat{S}_{TDG}$ = US$25 ($n$ = 158) and in the dictator game $\hat{S}_{DG}$ = US$12 ($n$ = 192) on average.[11] The summary statistics are presented in Tables B2 and B3 in the Appendix.

We first note that the mean amounts people are willing to send and return in our one-shot game scenarios are surprisingly close to the standard results in investment and dictator game experiments despite the fact that our design substantially differs from earlier designs. Camerer (2003) reports in his survey of experimental results that trustors typically send about half of their endowment (0.55 of their endowment in our case) and trustees return about the amount sent (1.1 times the amount sent in our case). Using a similar subject pool as ours, CEOs, Fehr and List (2004) find that their trustors send 0.59 of their endowment and return 1.3 times the amount sent.

Dictators generally send about 0.2 of their endowment (0.12 in our case) (Camerer 2003). The triple dictator game has not been studied widely. Ashraf et al. (2006) find that subjects send 0.24 of their endowment on

average (0.25 in our case). There are no significant differences in the amounts various demographic groups are willing to send in the dictator game scenario in our study. While women are willing to send slightly more, this difference is not significant. All demographic groups are willing to send more in the triple dictator game than in the dictator game. However, the increase is small for women and not significant. Men respond significantly more strongly to the efficiency gains and/or the decrease in the relative price of giving than women. This reversal between the dictator and the triple dictator game corresponds to the findings by Andreoni and Vesterlund (2001).

4.1 Institutions

Trustors' amounts sent and expected amounts returned and trustees' amounts returned and expected amounts sent are significantly[12] larger in all game scenarios with institutional constraints than in the one-shot game (see Tables B2 and B3). Given trustees' expectations of amounts sent, there are only small differences in proportions returned out of the expected amounts sent between the treatments. People maximally return 1.47 times the amount expected to be sent with pre-play communication and minimally 1.37 times the amount expected to be sent with the post-play punishment option (sig.).[13]

In most games, women are willing to trust less and expect to get back less than men (sig. but in post-communication treatment). Whites often trust and expect back more than non-whites (non-significant in one-shot and repeated games). There are no differences between Americans and non-Americans. Women are significantly less willing to reward trust and expect to be sent less than men in all treatment conditions. There are generally no significant differences in the willingness to be trustworthy (sig. in one-shot game) or expectations to be sent between whites and non-whites. Americans are significantly less willing to return money in the one-shot and in the punishment treatments. In the one-shot game scenario, non-Americans are significantly more optimistic than Americans.[14]

The summary statistics suggest that the amounts people are willing to send and return might be larger in the institutionally richer settings because trustors and trustees expect their counterparts to positively respond to the institutions and send and return more than in the one-shot game scenario. We now more precisely analyze the channels through which institutions affect trust and trustworthiness behavior.

4.2 Beliefs and Preferences

The first-stage regression results for expectations of trustworthiness are presented in Table B4. Column 1 shows that trustors become more optimistic about returns when institutional constraints are available. The order in which the games are played does not play any role. Women are less optimistic than men. In column 2, we include social orientation and find – unexpectedly – that the more socially oriented expect back higher returns. This might be due to a 'false consensus effect' where people base their expectations about others' behavior on their own type (Dawes 1989). In column 3, we also include the interaction variables between institutions and social orientation. Generally, social orientation matters more for beliefs in the one-shot game than in the institutionally constrained games, although some of the differences are not significant.

The second-stage regression results for trust are presented in Table B5. Column 1 suggests a strong relationship between institutions and trust: trustors send significantly more with institutional constraints than in the anonymous one-shot game. Women send less than men. The order in which the games are played is not relevant. Column 2 shows that the amounts trustors are willing to send in the triple dictator game are significantly related to the amounts sent in the trust game. However, a trustor's unconditional social orientation plays a substantially smaller role with institutional constraints than in the anonymous one-shot game (column 3). Social orientation completely loses its economic and statistical significance in the punishment treatment.

Figure 1 illustrates this graphically. There is no significant relationship between social orientation and trust in the punishment treatment because selfish people who give nothing in the triple dictator game send more and the more socially oriented send less in the punishment than in the anonymous one-shot game.

Institutions and the differences in the relevance of social orientation remain important when controlling for expectations of return in column 4. In column 5, we substitute expectations by their predicted values based on the first-stage regression (Table B4, column 3). As institutions and social orientation also affect behavior by changing expectations, the size of most coefficients decreases. However, institutions still exhibit a direct influence on behavior and also affect preferences in most cases. Notably, social orientation still matters in the anonymous one-shot game scenario but loses all its importance in the institutionally enriched game scenarios.

Our data suggest that institutional constraints decrease intrinsically motivated trust. Compared to the anonymous one-shot game, unconditional

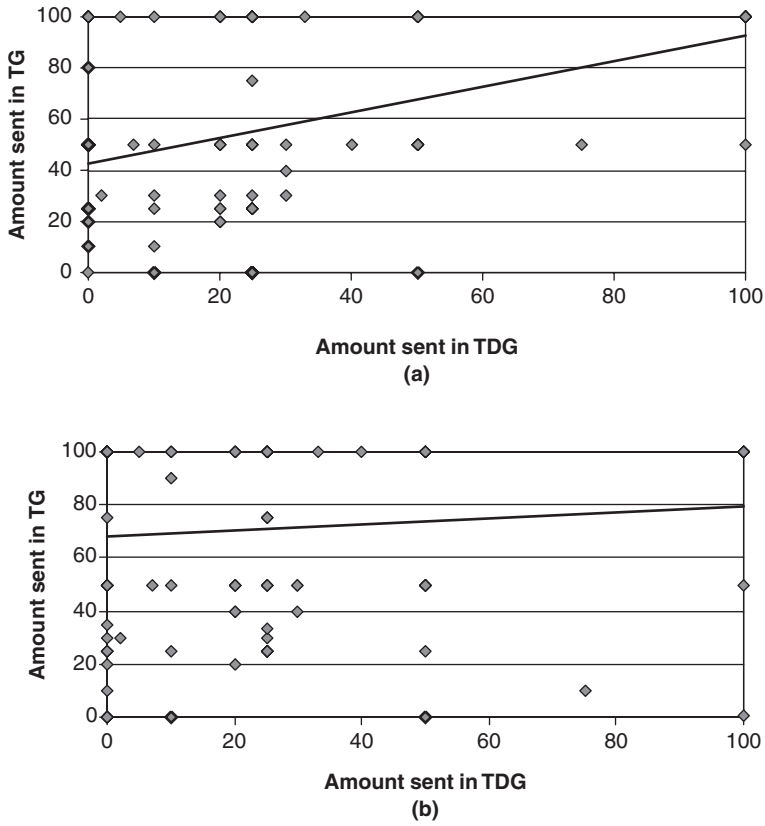


Figure 1. Trust and triple dictator game scenario behavior: (a) anonymous one-shot; (b) punishment

social orientation plays a significantly smaller, if any, role for willingness to trust in all institutionally constrained conditions. Institutional constraints crowd out the role unconditional social orientation plays for trust, supporting the *crowding hypothesis*.

Table B6 presents the results for the first-stage regressions on expectations of trust. Column 1 shows that trustees are more optimistic about the amounts of trust trustors offer with institutional constraints. Column 2 includes social orientation, which again is positively related to expectations. As before, women are more pessimistic than men. Column 3 suggests that social orientation plays a more important role for beliefs in the one-shot than in the institutionally constrained treatments.

Table B7 presents the results for trustworthiness. Column 1 shows that trustees return more when institutional constraints are available than in the anonymous one-shot game. Women return less than men. Order does not play any role. Column 2 suggests that unconditional social orientation is irrelevant for trustworthiness. However, in Column 3, we find that amounts sent in the DG are significantly less related to amounts returned in the institutionally constrained treatments than in the one-shot game. In fact, there is a significant negative relationship between unconditional social orientation and trustworthiness in the punishment condition.

Figure 2 illustrates that there is hardly any connection between what people are willing to send in the dictator game and what they are willing to return in the anonymous one-shot trust game. With punishment, the selfish tend to increase the amounts returned in comparison to the more socially oriented who tend to decrease the amounts returned compared to the anonymous one-shot game.

In column 4, we add expectations of amounts sent. They substantially change the picture. Institutions and social orientation lose their significance.[15] As institutions are not related to trustworthiness when controlling for the amounts trustees expect to be sent, we examine whether institutions may serve as an instrument for expectations of trust. Column 5 presents the IV regression results. They look very similar to column 4: Women return about US$12 less than men and trustees return about 1.5 times the amount they expect to be sent to them.

Our data suggest that institutions increasing the costs of betrayal (or the benefits of trustworthiness) hardly exhibit an independent influence on trustees' willingness to be trustworthy. Generally, institutions affect beliefs and trustees respond to those beliefs. The more they expect trustors to send to them, the more they are willing to return, suggesting reciprocity (Rabin 1993). Social orientation generally is not related to trustworthiness – although there is some indication for a negative relationship, particularly with punishment. All in all, we cannot reject the *null hypothesis* for trustworthiness: Institutional constraints mainly make trustees more optimistic about amounts sent. Given these expectations, they return more when constrained by institutions.

5. Conclusion

Institutions making trust betrayal more costly increase trustors' optimism about trustees' trustworthiness but at the same time, decrease their

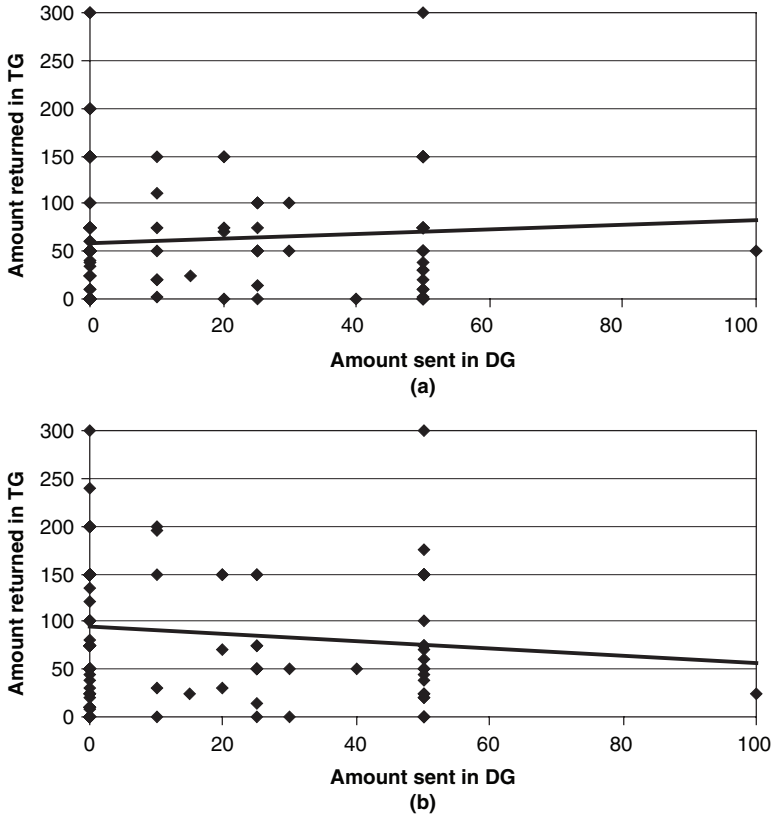


Figure 2. Trustworthiness and dictator game scenario behavior: (a) anonymous one-shot; (b) punishment

intrinsic willingness to assume losses by trusting. We measure a trustor's willingness to give up money for someone else's sake in dictator game scenarios. Such social orientation plays an important role in the anonymous one-shot trust game scenario but becomes irrelevant when trustee behavior is institutionally constrained. In contrast, our institutional constraints do not affect our trustees' willingness to reciprocate trust. They make trustees more optimistic about trustors' trust and only affect behavior through these expectations.

We conclude that institutions *may* affect behavior not only by changing beliefs but also by influencing motivation. Some intrinsic motivators

seem to be more robust than others. For example, our results suggest that in contrast to unconditional social orientation, reciprocity seems to be a more stable internalized norm. Our trustees are willing to return about the same fraction out of the amount they expect to receive, independent of the institutional environment. Gouldner (1960: 171), one of the first to point out the importance of this motivator, for example, wrote: 'A norm of reciprocity is, I suspect, no less universal and important an element of culture than the incest taboo.'

Overall, efficiency increases in our experiments when institutions make betrayal more costly. Trustors are willing to send more and trustees to return more than in an anonymous one-shot environment. Superficially, the institutional arrangements thus have achieved their purpose. We may only care about outcomes but not about why outcomes came to be. However, we may be concerned about the longer-term consequences for society if certain motivators lose in importance. In her book on *Trust in Society*, Cook (2001: xxvii), for example, points out: 'Even the law is a blunt instrument that cannot efficiently produce the kind of social order that comes from the existence of trusting relations in a group or society.'

Evolutionary theories of cultural transmission suggest that the crowding processes described here are self-enforcing. People copy the preferences of the successful and respond to the frequency with which certain preferences are held (Boyd and Richerson 1985; Bowles 1998; Bohnet et al. 2001). These dynamics can further be enhanced by selection effects. Environments characterized by external contract enforcement devices do not attract the intrinsically motivated 'good-doers' as they do not reward intrinsic social orientation. Politics will not attract particularly virtuous politicians if it treats the socially oriented identically to the selfish.

Organizations and societies differ in the frequency with which interactions are governed by formal contracts and punishment as opposed to intrinsic trust. Field examples of interactions based on trust include Bernstein's (1992) analysis of the international diamond market and Siamwalla's (1978) discussion of the raw rubber market in Asia. Ostrom et al. (1994) discuss a large number of social dilemma situations, which people were able to overcome relying on trust and informal guarantees. More formal enforcement mechanisms often interfered with these trust-based interactions.

Institutions, thus, serve two purposes: the traditionally discussed purpose of keeping people from being too selfish and not pursuing the principals' interests or the common good. Constitutions are for 'knaves' (Brennan and Buchanan 1983) – but as Frey (1997) has pointed out: 'A

constitution for knaves crowds out civic virtues.' The second purpose must be to protect and nurture individuals' intrinsic motivation, their civic virtue, their willingness to trust and be trustworthy absent material incentives to do so.

NOTES

1. Sequential games examining trust and trustworthiness include, e.g., the binary-choice trust game (Camerer and Weigelt 1988; Kreps 1990) and the investment game (Berg et al. 1995). A number of simultaneous-choice games, e.g., the public goods game, have also been used to measure trust (see for a discussion of the various approaches, Bohnet and Croson 2004).
2. The effects of extrinsic incentives on intrinsic motivation were first explored in psychology (for a review, see Deci et al. 1999) and then incorporated into other social sciences (Frey 1997). Recent empirical tests include Frey and Oberholzer-Gee (1997) examining the effect of compensation on people's willingness to accept a NIMBY project, Gneezy and Rustichini (2000) examining the effect of fines on parents' willingness to pick up their children from daycare on time, and Fehr and Gächter (2002) examining the effect of punishment on voluntary cooperation. Frey and Jegen (2001) provide a survey of the empirical literature.
3. Using different theoretical frameworks, attribution theory and repeated game theory respectively, Malhotra and Murnighan (2002) and Miller (2004) come to similar conclusions.
4. For recent surveys and collections of experimental work on trust, see, for example, Camerer (2003), Ostrom and Walker (2003), and Bohnet and Croson (2004). Others have relied on survey data to measure trust, e.g. Putnam (1993), Fukuyama (1995), Brehm and Rahn (1997), Knack and Keefer (1997), Rahn and Transue (1998), Inglehart (1999), Robinson and Jackson (2001) and Alesina and La Ferrara (2002). Glaeser et al. (2000), Fehr et al. (2002), Bellemare and Kroeger (2003) and Ashraf et al. (2006) compare experimental and survey trust data.
5. 'Many people think within-subject analysis is the only proper analysis in choice experiments, because expected utility requires consistency of individual preferences' (Camerer 1995: 633).
6. The hypothetical games were conducted like a 'standard experiment': The instructions for the trust games, the (triple) dictator game, and the questionnaire collecting basic demographic data were distributed one at a time. After subjects had read the instructions, the experimenter read the instructions aloud. Participants were then invited to ask questions in private (hardly any questions were asked). Subjects then had to write down their decision for game 1, fold the decision form and put it into a box. The decision forms for the remainder of the games were distributed, one at a time, following the same procedure as above. Sample instructions can be found in Appendix A.
7. Only very few studies are based on non-student subject pools. In a recent paper, Fehr and List (2004) run anonymous one-shot trust games with CEOs in Costa Rica. We will compare our findings for the one-shot game with theirs.

8. Note that in the last session, we accidentally distributed trustee roles to all 36 participants. Subjects, and in effect, the experimenter, were not aware of this until the response forms were collected. Thus, in the end, it turned out that we had more people in the role of the trustee ($n = 194$) than in the role of the trustor ($n = 158$). As people are not matched according to their decisions in survey experiments, this is not a problem.
9. Eliciting beliefs in addition to having subjects make action choices has been found to increase 'rationality'. Subjects are more likely to take their counterpart's incentives into account when beliefs are elicited than when they make action choices only (Croson 2000).
10. Models incorporating incomplete information about trustees' preferences and/or rationality allow for positive amounts sent and returned also in finitely repeated games (Kreps et al. 1982).
11. In our control group, in which subjects participated in one game scenario only, trustors are willing to send US$51 ($n = 20$) and trustees to return US$54 ($n = 20$). Dictators are willing to send US$26 in the triple dictator game ($n = 19$) and US$19 in the dictator game ($n = 20$) on average. The amounts sent and returned (and the expectations thereof) in the control scenarios do not significantly differ from the amounts people are willing to send and return (and the expectations thereof) in our main experiment.
12. We report a result as significant if $p < 0.05$. For all comparisons, we use a non-parametric test for differences in means, the Mann–Whitney U-test.
13. The results in our control experiment are: Trust amount sent with pre-play communication: US$75 ($n = 15$), with post-play communication: US$66 ($n = 14$), with post-play punishment option: US$65 ($n = 14$), with repetition US$80 ($n = 15$). Trustworthiness amount returned with pre-play communication: US$108 ($n = 13$), with post-play communication: US$98 ($n = 14$), with post-play punishment option: US$99 ($n = 15$), with repetition US$115 ($n = 15$). We make three observations: (1) The amounts people are willing to send and return are remarkably similar in the control experiment and in our main experiment. (2) The pattern across institutional contexts is remarkably similar in both experiments: People are willing to send and return less in the one-shot game than in the institutionally richer games. (3) There is somewhat more variation in behavior between the institutionally richer games in the control than in the main experiment, suggesting that people differentiate more between different institutional settings when being confronted with one at a time than when participating in all sequentially.
14. See Yamagishi and Yamagishi (1994), Buchan et al. (2003), Eckel and Grossman (2004), Cardenas and Carpenter (2005) and Greig and Bohnet (2005) for a discussion of gender, race and cultural differences in trust games. Generally, the evidence on demographic characteristics is mixed.
15. Note that interactions between demographic variables and expectations of amounts sent were not significant.

REFERENCES

Alesina, A. and E. La Ferrara. 2002. 'Who Trusts Others?' *Journal of Public Economics* 85: 207–34.

Andreoni, J. 1990. 'Impure altruism and donations to public goods: A theory of warm-glow giving?' *Economic Journal* 100: 464–77.

Andreoni, J. and J. Miller. 2002. 'Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism.' *Econometrica* 70: 737–53.
Andreoni, J. and L. Vesterlund. 2001. 'Which is the Fair Sex? Gender Differences in Altruism.' *Quarterly Journal of Economics* 116: 293–312.
Ashraf, N., I. Bohnet and N. Piankov. 2006. 'Decomposing Trust and Trustworthiness.' *Experimental Economics* 9: 193–208.
Berg, J., J. Dickhaut and K. A. McCabe. 1995. 'Trust, Reciprocity, and Social History.' *Games and Economic Behavior* 10: 290–307.
Bellemare, C. and S. Kröger. 2003. 'On Representative Trust.' Discussion paper #47, Center for Economic Research, Tilburg University.
Bernstein, L. 1992. 'Opting out of the legal system: Extralegal contractual relations in the diamond industry.' *Journal of Legal Studies* 21: 115–57.
Bohnet, I. and R. Croson. 2004. 'Trust and Trustworthiness.' *Journal of Economic Behavior and Organization* 55: 443–5.
Bohnet, I. and S. Huck. 2004. 'Repetition and Reputation: Implications for Trust and Trustworthiness When Institutions Change.' *American Economic Review* 94: 362–6.
Bohnet, I., B. S. Frey and S. Huck. 2001. 'More Order With Less Law: On Contract Enforcement, Trust and Crowding.' *American Political Science Review* 89: 335–9.
Bolton, G. and A. Ockenfels. 2000. 'A Theory of Equity, Reciprocity and Competition.' *American Economic Review* 90: 166–93.
Bowles, S. 1998. 'Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions.' *Journal of Economic Literature* 35: 75–111.
Boyd, R. and P. J. Richerson. 1985. *Culture and the Evolutionary Process.* Chicago, IL: University of Chicago Press.
Brehm, J. and W. M. Rahn. 1997. 'Individual-level Evidence for the Causes and Consequences of Social Capital.' *American Journal of Political Science* 41: 999–1023.
Brennan, G. and J. M. Buchanan. 1983. 'Predictive Power and the Choice Among Regimes.' *Economic Journal* 93: 89–105.
Buchan, N., R. Croson and S. Solnick. 2003. 'Trust and Gender: An Examination of Behavior, Biases, and Beliefs in the Investment Game.' Working paper, Wharton School, University of Pennsylvania.
Camerer, C. 2003. *Behavioral Game Theory.* Princeton, NJ: Princeton University Press.
Camerer, C. 1995. 'Individual Decision Making.' In *The Handbook of Experimental Economics*, ed. J. Kagel and A. Roth, pp. 587–703. Princeton, NJ: Princeton University Press.
Camerer, C. and K. Weigelt. 1988. 'Experimental Tests of a Sequential Equilibrium Reputation Model.' *Econometrica* 56: 1–36.
Cardenas, J. C. and J. Carpenter. 2005. 'Three Themes on Field Experiments and Economic Development.' Working paper, Middlebury College.
Coleman, J. S. 1990. *Foundations of Social Theory.* Cambridge, MA: Harvard University Press.
Cook, K. (ed.). 2001. *Trust in Society.* New York: Russell Sage Foundation.
Cox, J.C. 2004. 'How To Identify Trust and Reciprocity.' *Games and Economic Behavior* 46: 260–81.
Crawford, S. and E. Ostrom. 1995. 'A Grammar of Institutions.' *American Political Science Review* 89: 582–601.
Croson, R. 2000. 'Thinking Like a Game Theorist: Factors Affecting the Frequency of Equilibrium Play.' *Journal of Economic Behavior and Organization* 41: 299–314.
Dawes, R. M. 1989. 'Statistical Criteria for Establishing a Truly False Consensus Effect.' *Journal of Experimental Social Psychology* 25: 1–17.

Deci, E., R. Koestner and R. Ryan. 1999. 'A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation.' *Psychological Bulletin* 125: 627–68.

Duch, R. M. and H. D. Palmer. 2004. 'It's Not Whether You Win or Lose, But How You Play the Game: Self-interest, Social Justice, and Mass Attitudes toward Market Transition.' *American Political Science Review* 98: 437–52.

Dufwenberg, M. and U. Gneezy. 2000. 'Measuring Beliefs in an Experimental Lost Wallet Game.' *Games and Economic Behavior* 30: 163–82.

Eckel, C. C. and P. J. Grossman. Forthcoming. 'Differences in the Economic Decisions of Men and Women: Experimental Evidence.' In *Handbook of Results in Experimental Economics*, eds C. Plott and V.L. Smith. Amsterdam: North-Holland.

Engle-Warnick, J. and R. L. Slonim. 2004. 'The Evolution of Strategies in a Repeated Trust Game.' *Journal of Economic Behavior and Organization* 55: 553–74.

Farrell, J. and M. Rabin. 1996. 'Cheap Talk.' *Journal of Economic Perspectives* 10: 103–18.

Fehr, E. and K. Schmidt. 1999. 'A Theory of Fairness, Competition and Cooperation.' *Quarterly Journal of Economics* 114: 817–68.

Fehr, E. and S. Gächter. 2002. 'Do Incentive Contracts Crowd Out Voluntary Cooperation?' Working paper, Institute for Empirical Economic Research in Economics, University of Zürich.

Fehr, E. and J. A. List. 2004. 'The Hidden Costs and Returns of Incentives.' *Journal of the European Economic Association* 2: 743–71.

Fehr, E., U. Fischbacher, B. von Rosenbladt, J. Schupp and G. G. Wagner. 2002. 'A Nation-wide Laboratory: Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Surveys.' *Schmollers Jahrbuch* 122: 519–42.

Frey, B. S. 1997. *Not Just For the Money. An Economic Theory of Personal Motivation.* Cheltenham: Edward Elgar.

Frey, B. S. and R. Jegen. 2001. 'Motivation crowding theory: A survey of empirical evidence.' *Journal of Economic Surveys* 15: 589–611.

Frey, B.S. and F. Oberholzer-Gee. 1997. 'The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding Out.' *American Economic Review* 87: 746–55.

Fukuyama, F. 1995. *Trust: Social Virtues and the Creation of Prosperity.* New York: Free Press.

Gächter, S., E. Fehr and C. Kment. 1996. 'Does Social Exchange Increase Voluntary Cooperation?' *Kyklos* 49: 541–54.

Gambetta, D. 1988. 'Can We Trust Trust?' In *Trust: Making and Breaking Cooporative Relations*, ed. D. Gambetta, pp. 213–37. Oxford: Basil Blackwell.

Glaeser, E. L., D. I. Laibson, J. A. Scheinkman and C. L. Soutter. 2000. 'Measuring Trust.' *Quarterly Journal of Economics* CXV: 811–46.

Gneezy, U. and A. Rustichini. 2000. 'A fine is a price.' *Journal of Legal Studies* 29: 1–17.

Gouldner, A. 1960. 'The Norm of Reciprocity.' *American Sociological Review* 25: 161–78.

Granovetter, M. 1992. 'The Sociological and Economic Approaches to Labour Market Analysis: A Social Structural View.' In *The Sociology of Economic Life*, eds M. Granovetter and R. Swedberg, pp. 233–63. Boulder, CO: Westview Press.

Greig, F. and I. Bohnet. 2005. 'Is There Reciprocity in a Reciprocal-exchange Economy? Evidence from a Nairobi Slum.' Working paper, Kennedy School of Government, Harvard University.

Hardin, R. 2002. *Trust and Trustworthiness.* New York: Russell Sage.

Hobbes, T. 1960. *Leviathan.* Oxford: Basil Blackwell.

Inglehart, R. 1999. 'Trust, Well-being and Democracy.' In *Democracy and Trust*, ed. M. E. Warren, pp. 88–120. Cambridge: Cambridge University Press.
Kahneman, D., J. Knetsch and R. Thaler. 1986. 'Fairness as Constraint on Profit Seeking: Entitlements in the Market.' *American Economic Review* 76: 728–41.
Knack, S. 2002. 'Social capital and the quality of government: Evidence from the American states.' *American Journal of Political Science* 46: 772–85.
Knack, S. and P. Keefer. 1997. 'Does Social Capital Have an Economic Payoff? A Cross-Country Investigation.' *Quarterly Journal of Economics* 112: 1251–88.
Kramer, R. 1999. 'Trust and Distrust in Organizations: Emerging Perspectives, Enduring Questions.' *Annual Review of Psychology* 50: 569–98.
Kramer, R. and T. R. Tyler (eds). 1996. *Trust in Organizations.* Thousand Oaks: Sage.
Kreps, D. 1990. 'Corporate Culture and Economic Theory.' In *Perspectives on Positive Political Economy*, eds J. E. Alt and K. A. Shepsle, pp. 90–143. Cambridge: Cambridge University Press.
Kreps, D. P., J. Milgrom, J. Roberts and R. Wilson. 1982. 'Rational Cooperation in the Finitely Repeated Prisoners' Dilemma.' *Journal of Economic Theory* 27: 245–52.
LaPorta, R., F. Lopez-de-Silanes, A. Shleifer and R.W. Vishny. 1997. 'Trust in Large Organizations.' *American Economic Review* 87: 333–38.
Malhotra, D. and J. K. Murnighan. 2002. 'The Effects of Formal and Informal Contracts on Interpersonal Trust.' *Administrative Science Quarterly* 47: 534–59.
Mansbridge, J. 1999. 'Altruistic Trust.' In *Democracy and Trust*, ed. M. E. Warren, pp. 290–309. Cambridge: Cambridge University Press.
Miller, G. J. 2004. 'Monitoring, Rules, and the Control Paradox: Can the Good Soldier Švejk Be Trusted?' In *Trust and Distrust in Organizations*, eds R. M. Kramer and K. S. Cook, pp. 99–126. New York: Russell Sage.
Nye, J. S., P. D. Zelikow and D. King (eds). 1997. *Why People Don't Trust Government.* Cambridge, MA: Harvard University Press.
Ostrom, E. 1998. 'A Behavioral Approach to the Rational Choice Theory of Collective Action.' *American Political Science Review* 92: 1–22.
Ostrom, E. and J. Walker (eds). 2003. *Trust and Reciprocity.* New York: Russell Sage.
Ostrom, E., R. Gardner and J. Walker. 1994. *Rules, Games and Common-Pool Resources*. Ann Arbor, MI: University of Michigan Press.
Ostrom, E., J. Walker and R. Gardner. 1992. 'Covenants With and Without a Sword: Self-governance is Possible.' *American Political Science Review* 86: 404–17.
Putnam, R. D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy.* Princeton, NJ: Princeton University Press.
Putnam, R. D. 2000. *Bowling Alone. The Collapse and Revival of American Community.* New York: Simon & Schuster.
Pindyck, R. S. and D. L. Rubinfeld. 1991. *Econometric Models & Economic Forecasts*. New York: McGraw-Hill.
Rabin, M. 1993. 'Incorporating Fairness into Game Theory and Economics.' *American Economic Review* 83: 1281–302.
Rahn, W. M. and J. E. Transue. 1998. 'Social Trust and Value Change: The Decline of Social Capital in American Youth, 1976–1995.' *Political Psychology* 19: 545–65.
Robinson, R. and E. Jackson. 2001. 'Is Trust in Others Declining in America? An Age-Period-Cohort Analysis.' *Social Science Research* 30: 117–45.
Rosenfeld, R., S. F. Messner and E. P. Baumer. 2001. 'Social Capital and Homicide.' *Social Forces* 80: 283–309.

Rousseau, D., S. Sitkin, R. Burt and C. Camerer. 1998. 'Not So Different After All: A Cross-discipline View of Trust.' *Academy of Management Review* 23: 393–404.
Sally, D. 1995. 'Conversation and Cooperation in Social Dilemmas.' *Rationality and Society* 7: 58–92.
Siamwalla, A. 1978. 'Farmers and Middlemen: Aspects of Agricultural Marketing in Thailand.' *Economic Bulletin for Asia and the Pacific* 39: 38–50.
Sitkin, S. B. and N. L. Roth. 1993. 'Explaining the Limited Effectiveness of Legalistic "Remedies" for Trust/Distrust.' *Organization Science* 4: 367–92.
Yamagishi, T. and M. Yamagishi. 1994. 'Trust and Commitment in the United States and Japan.' *Motivation and Emotion* 18: 129–66.
Zak, P. and S. Knack. 2001. 'Trust and Growth.' *Economic Journal* 111: 295–321.

Appendix A: Sample Instructions

First Movers

Thank you for participating in our research. **You have been randomly assigned to the role of #1 in this study**. The study consists of three sections. The instructions will be on 8 different pages. Please keep page 1 for your reference.

Section I: The questions in this section refer to variations on one basic scenario.

Basic scenario:

In this scenario, there are two participants who are randomly assigned the roles of #1 and #2. **#1 is given $100** at the beginning, and faces the first decision: whether or not to pass along any or all of the $100 to #2. #1 may send no money, all the money, or any amount in-between. **Any money sent by #1 to #2 will be automatically tripled before it is received by #2.**

> For example, if #1 sends $0 then #2 receives $0; if #1 sends $100 then #2 receives $300; if #1 sends $50 then #2 receives $150; etc.

If #1 chooses to send no money, then the game ends and #1 leaves with $100 and #2 leaves with nothing.

If #1 chooses to send some amount of money, then #2 faces the next decision: how much money to send back to #1. #2 may send no money, all the money, or any amount in-between. **#1 receives exactly the amount of money sent by #2**. Each participant leaves with the money he/she is holding at the end of this interaction.

Note: Both participants are aware of all the rules, the initial amount of money received by the other, all money-decisions made by the other, and that all decisions are hypothetical.

– new page, subjects keep this page –

1. Assume the scenario is a one-shot occurrence with complete anonymity.
   (a) As #1 starting with $100, how much money (if any) would you send to #2? $_____
   (b) How much money (if any) do you think #2 would return to you? $_____

– new page, subjects hand in this page –

2. Assume the scenario is a one-shot occurrence and that #1 and #2 must meet for 5 minutes before #1 decides how much money to send to #2 and before #2 decides how much money to return to #1.
   (a) As #1 starting with $100, how much money (if any) would you send to #2? $_____
   (b) How much money (if any) do you think #2 would return to you? $_____

– new page, subjects hand in this page –

3. Assume the scenario is a one-shot occurrence and that #1 and #2 must meet for 5 minutes after both #1 and #2 have decided how much money to send each other and both have learned about their counterpart's choices.
   (a) As #1 starting with $100, how much money (if any) would you send to #2? $_____
   (b) How much money (if any) do you think #2 would return to you? $_____

– new page, subjects hand in this page –

4. This scenario is a one-shot occurrence and a third stage is added, such that after both #1 and #2 have decided how much money to send each other, #1 has the option to decrease the amount of money that #2 holds by 50%. If #1 chooses this option, #1 does not gain or lose any money; the money is simply taken away from #2.
   (a) As #1 starting with $100, how much money (if any) would you send to #2? $_____
   (b) How much money (if any) do you think #2 would return to you? $_____

– new page, subjects hand in this page –

5. Assume the scenario will be repeated 5 times with the same partner with complete anonymity. Also assume that #1 starts with $20 instead of $100 at the beginning of each round (i.e. a maximum of $20 can be sent to #2 in each round). The total amount of money available over all five rounds is $100.

   (a) As #1 having $20 available in each round, how much money (if any) would you send to #2 on average in each round? $_____
   (b) How much money (if any) do you think #2 would return to you on average in each round? $_____

– new page, subjects hand in this page –

Section II: In this section there is one hypothetical scenario.

6. In this scenario, there are two participants who are randomly assigned the roles of #1 and #2. **#1 is given $100** at the beginning, and faces the following decision: whether or not to pass along any or all of the $100 to #2. #1 may send no money, all the money, or any amount in-between. **Any money sent by #1 to #2 will be automatically tripled before it is received by #2.**

   For example, if #1 sends $0 then #2 receives $0; if #1 sends $100 then #2 receives $300; if #1 sends $50 then #2 receives $150; etc.

The game ends after #1 has made a decision.

As #1 starting with $100, how much money (if any) would you send to #2? $_____

– new page, subjects hand in this page –

Section III: Basic demographic characteristics.

7. What country are you from (where have you lived most of your life)?

   _____

8. What is the ethnic group with which you most identify?

   African or African American ☐ Asian or Pacific Islander ☐ Caucasian ☐ Hispanic ☐

   Native American ☐ Other ☐ (Please indicate_____)

9. What is your gender?

   male ☐ female ☐

.....

Second Movers

Thank you for participating in our research. **You have been randomly assigned to the role of #2 in this study.** The study consists of three sections. The instructions will be on 8 different pages. Please keep page 1 for your reference.

Section I: The questions in this section refer to variations on one basic scenario.

Basic scenario:

In this scenario, there are two participants who are randomly assigned the roles of #1 and #2. **#1 is given $100** at the beginning, and faces the first decision: whether or not to pass along any or all of the $100 to #2. #1 may send no money, all the money, or any amount in-between. **Any money sent by #1 to #2 will be automatically tripled before it is received by #2.**

> For example, if #1 sends $0 then #2 receives $0; if #1 sends $100 then #2 receives $300; if #1 sends $50 then #2 receives $150; etc.

If #1 chooses to send no money, then the game ends and #1 leaves with $100 and #2 leaves with nothing.

If #1 chooses to send some amount of money, then #2 faces the next decision: how much money to send back to #1. #2 may send no money, all the money, or any amount in-between. **#1 receives exactly the amount of money sent by #2**. Each participant leaves with the money he/she is holding at the end of this interaction.

Note: Both participants are aware of all the rules, the initial amount of money received by the other, all money-decisions made by the other, and that all decisions are hypothetical.

– new page, subjects keep this page –

1. Assume the scenario is a one-shot occurrence with complete anonymity.

   (a) As #2, how much money (if any) of the $100 do you think #1 would send to you? $_____

   (b) Assuming #1 sends the amount of money you listed in Part (a) and you receive triple that amount, how much money would you send to #1? $_____

– new page, subjects hand in this page –

2. Assume the scenario is a one-shot occurrence and that #1 and #2 must meet for 5 minutes before #1 decides how much money to send to #2 and before #2 decides how much money to return to #1.

(a) As #2, how much money (if any) of the $100 do you think #1 would send to you? $_____
(b) Assuming #1 sends the amount of money you listed in Part (a) and you receive triple that amount, how much money would you send to #1? $_____

– new page, subjects hand in this page –

3. Assume the scenario is a one-shot occurrence and that #1 and #2 must meet for 5 minutes after both #1 and #2 have decided how much money to send each other and both have learned about their counterpart's choices.

(a) As #2, how much money (if any) of the $100 do you think #1 would send to you? $_____
(b) Assuming #1 sends the amount of money you listed in Part (a) and you receive triple that amount, how much money would you send to #1? $_____

– new page, subjects hand in this page –

4. This scenario is a one-shot occurrence and a third stage is added, such that after both #1 and #2 have decided how much money to send each other, #1 has the option to decrease the amount of money that #2 holds by 50%. If #1 chooses this option, #1 does not gain or lose any money; the money is simply taken away from #2.

(a) As #2, how much money (if any) of the $100 do you think #1 would send to you? $_____
(b) Assuming #1 sends the amount of money you listed in Part (a) and you receive triple that amount, how much money would you send to #1? $_____

– new page, subjects hand in this page –

5. Assume the scenario will be repeated 5 times with the same partner with complete anonymity. Also assume that #1 starts with $20 instead of $100 at the beginning of each round (i.e. a maximum of $20 can be sent to #2 in each round). The total amount of money available over all five rounds is $100.

(a) As #2, how much money (if any) of the $20 do you think #1 would send to you on average in each round? $_____

(b) Assuming #1 sends the amount of money you listed in Part (a) and you receive triple that amount, how much money would you send to #1 on average in each round? $_____

– new page, subjects hand in this page –

Section II: In this section there is one hypothetical scenario.

6. In this scenario, there are two participants who are randomly assigned the roles of #1 and #2. **#2 is given $100** at the beginning, and faces the following decision: whether or not to pass along any or all of the $100 to #1. #2 may send no money, all the money, or any amount in-between.

   For example, if #2 sends $0 then #1 receives $0; if #2 sends $100 then #1 receives $100; if #2 sends $50 then #1 receives $50; etc.

The game ends after #2 has made a decision.

As #2 starting with $100, how much money (if any) would you send to #1? $_____

– new page, subjects hand in this page –

Section III: Basic demographic characteristics

Identical for both players.

Appendix B

Table B1. Variable definitions

Concept	*Variable name*	*Variable description*	*Range of values*
Willingness to trust	Amount sent, $\hat{S}$	Amount trustor willing to send in trust game	US$0–100
Willingness to reward trust	Amount returned, $\hat{Y}$	Amount trustee willing to return in trust game	US$0–300
Social orientation for trust	TDG giving, $\hat{S}_{TDG}$	Amount subject willing to send in triple dictator game	US$0–100
Social orientation for trustworthiness	DG giving, $\hat{S}_{DG}$	Amount subject willing to send in dictator game	US$0–100
Expectations of trustworthiness	Expected back, $E(\hat{Y})$	Amount trustor expects to be sent back by trustee	US$0–300
Expectations of trust	Expected to be sent, $E(\hat{S})$	Amount trustee expects to be sent by trustor	US$0–100
Version of experiment	Version	Version: 0 if dictator game first 1 if trust games first	0 or 1

Table B2. Summary statistics: willingness to trust (amounts sent and expected back in US$)

	ALL			*MEN*			*WOMEN*			*WHITES*			*NON-WHITES*			*AMERICANS*			*NON-AMERICANS*		
	M	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*
Amount sent: one-shot	55.19	39.65	158	60.28	41.39	89	48.62	36.53	69	56.64	39.81	110	51.88	39.49	48	52.89	40.03	97	58.17	39.03	60
Amount sent: pre-communication	69.74	35.75	155	75.51	34.13	88	62.16	36.65	67	72.59	35.44	108	63.19	35.96	47	68.92	36.19	97	70.61	35.42	57
Amount sent: post-communication	68.78	36.47	152	72.97	34.85	86	63.33	38.05	66	71.95	36.30	105	61.70	36.21	47	67.21	36.81	95	70.89	36.17	56
Amount sent: punishment	70.69	35.95	150	75.21	35.85	86	64.61	35.45	64	74.88	33.89	104	61.196	38.95	46	71.26	35.68	95	69.13	36.83	54
Amount sent: repetition	68.42	37.65	146	74.29	35.86	84	60.48	38.82	62	71.44	37.15	101	61.67	38.29	45	67.53	37.14	93	69.42	38.97	52
Amount expected back: one-shot	70.1	62.25	150	79.89	67.48	88	56.21	51.34	62	72.30	61.91	102	65.42	63.36	48	69.22	60.65	90	70.08	64.80	59
Amount expected back: pre-communication	95.03	68.48	152	107.87	73.16	88	77.37	57.46	64	99.66	69.58	105	84.68	65.49	47	96.77	65.49	94	91.18	73.81	57

(continued)

Table B2. (continued)

	ALL			*MEN*			*WOMEN*			*WHITES*			*NON-WHITES*			*AMERICANS*			*NON-AMERICANS*		
	M	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*
Amount expect. back: post-communication	93.89	68.44	148	103.53	69.34	85	80.87	65.52	63	101.34	69.85	101	77.87	63.09	47	93.76	66.80	91	93.08	71.84	56
Amount expected back: punishment	95.78	68.69	148	107.74	72.71	86	79.19	59.34	62	100.99	67.64	102	84.24	70.32	46	94.46	62.96	93	97.05	78.39	54
Amount expected back: repetition	89.67	72.07	142	101.96	69.71	83	72.37	72.36	59	95.70	72.06	97	76.67	71.14	45	85.51	68.91	89	95.63	77.64	52
TDG giving	25.49	31.46	158	31.49	35.87	89	17.74	22.63	69	23.25	30.14	110	30.60	34.07	48	20.07	0.46	97	34.12	35.81	60

Table B3. Summary statistics: willingness to trust (amounts returned and expected to be sent in US$)

	ALL			*MEN*			*WOMEN*			*WHITES*			*NON-WHITES*			*AMERICANS*			*NON-AMERICANS*		
	M	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*
Amount returned: one-shot	61.02	63.62	182	70.02	68.24	117	44.82	50.92	65	66.19	64.26	116	51.92	61.9	66	53.29	62.27	117	74.92	64.15	65
Amount returned: pre-communication	111.06	65.35	193	121.44	63.84	125	91.99	64.24	68	116.07	61.56	124	102.07	71.26	69	110.91	63.82	124	111.35	68.50	69
Amount returned: post-communication	94.06	69.74	188	102.87	71.19	122	77.77	64.36	66	96.13	65.52	122	90.23	77.31	66	92.76	70.12	123	96.51	69.47	65
Amount returned: punishment	88.73	68.15	181	98.21	68.73	115	72.21	64.34	66	92.89	66.83	116	81.31	70.35	65	81.96	68.46	117	101.11	66.32	64
Amount returned: repetition	93.99	73.56	172	101.97	77.42	107	80.85	65.17	65	98.04	76.31	112	86.42	68.09	60	89.00	69.99	112	103.31	79.55	60
Amount expected to be sent: one-shot	41.76	35.75	193	45.12	36.95	125	35.58	32.79	68	42.84	36.78	125	39.77	33.95	68	37.39	35.39	125	49.78	35.26	68
Amount expected to be sent: pre-communication	75.45	34.14	192	78.36	31.69	125	70.01	37.96	67	78.31	31.62	124	70.24	38.01	68	76.11	35.30	123	74.28	32.18	69
Amount expected to be sent: post-communication	66.01	38.22	189	67.68	34.39	123	62.90	44.61	66	66.17	35.26	122	65.72	43.38	67	65.76	40.79	123	66.48	33.20	66

(continued)

Table B3. (continued)

	ALL			*MEN*			*WOMEN*			*WHITES*			*NON-WHITES*			*AMERICANS*			*NON-AMERICANS*		
	M	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*	*M*	SD	*N*
Amount expected to be sent: punishment	64.83	35.33	182	67.82	35.46	117	59.44	34.71	65	65.00	36.57	117	64.51	33.26	65	62.82	36.65	118	68.52	32.72	64
Amount expected to be sent: repetition	64.55	36.30	176	66.44	37.25	111	61.31	34.67	65	65.00	36.85	114	63.71	35.55	62	63.30	37.10	115	66.89	34.93	61
DG giving	11.90	20.19	192	10.32	20.27	124	14.78	19.88	68	11.16	19.05	125	13.28	22.25	67	11.37	19.94	124	12.87	20.76	68

Table B4. Determinants of expectations of trustworthiness, E(Ŷ)

	(1)	*(2)*	*(3)*
Pre-communication	24.988***	25.178***	32.672***
	(4.679)	(4.689)	(6.602)
Post-communication	24.060***	24.265***	27.883***
	(4.819)	(4.825)	(6.602)
Punishment	25.635***	26.025***	37.657***
	(4.814)	(4.811)	(6.311)
Repetition	19.395***	19.851***	22.288***
	(5.996)	(6.012)	(7.432)
Version (trust first)	10.501	10.544	10.419
	(9.216)	(9.089)	(9.116)
Female	–24.149**	–20.034**	–20.050**
	(9.389)	(9.467)	(9.492)
White	12.821	15.804	15.713
	(9.865)	(9.521)	(9.552)
American	0.783	4.806	4.755
	(9.939)	(9.873)	(9.900)
Social Orientation TDG		0.361 ***	0.552***
		(0.133)	(0.143)
TDG × pre-communication			–0.287**
			(0.143)
TDG × post-communication			–0.135
			(0.135)
TDG × punishment			–0.356**
			(0.146)
TDG × repetition			–0.086
			(0.150)
Constant	65.586***	49.880***	44.991***
	(12.942)	(14.101)	(14.089)
R^2	0.071	0.100	0.101
N of clusters	155	155	155
Observations	735	735	735

Notes: ***significant at 1% level; **significant at 5% level; *significant at 10%, corrected standard errors in parentheses.

Table B5. Determinants of willingness to trust, $\hat{S}$

	(1)	*(2)*	*(3)*	*(4)*	*(5) predicted expectations*
Pre-communication	13.228***	13.312***	20.919***	6.105***	3.709***
	(2.561)	(2.578)	(3.425)	(1.883)	(1.183)
Post-communication	10.981***	11.051***	16.531***	5.579***	1.931
	(2.592)	(2.610)	(3.582)	(1.887)	(1.245)
Punishment	11.917***	11.994***	23.668***	7.393***	2.261**
	(3.048)	(3.069)	(3.748)	(2.187)	(1.084)
Repetition	13.488***	13.292***	16.912***	6.716**	5.144**
	(2.992)	(2.956)	(4.025)	(2.988)	(2.261)
Version (trust first)	7.301	7.858	7.869	3.228	
	(5.158)	(5.004)	(5.015)	(2.549)	
Female	–9.938*	–7.224	–7.239	4.734*	2.008
	(5.454)	(5.394)	(5.407)	(2.607)	(3.087)
White	6.633	9.075*	9.141	6.227**	1.791
	(5.736)	(5.558)	(5.569)	(2.569)	(3.956)
American	0.242	4.068	4.061	1.467	1.853
	(5.566)	(5.380)	(5.391)	(2.699)	(3.065)
Social orientation		0.274***	0.499***	0.234***	0.162**
TDG		(0.069)	(0.073)	(0.049)	(0.068)
TDG × pre-communication			–0.299***	–0.118**	–0.093*
			(0.077)	(0.046)	(0.049)
TDG × post-communication			–0.215***	–0.095**	–0.079*
			(0.060)	(0.039)	(0.041)
TDG × punishment			–0.459***	–0.159**	–0.131*
			(0.097)	(0.067)	(0.071)
TDG × repetition			–0.139**	–0.077	–0.124**
			(0.067)	(0.055)	(0.058)
Expectations of amounts returned				0.425***	0.461***
				(0.024)	(0.098)
Constant	50.663***	38.592***	32.837***	12.834***	15.605**
	(7.027)	(7.907)	(7.965)	(3.673)	(6.715)
R^2	0.056	0.107	0.123	0.718	0.107
N of clusters	158	157	157	155	157
Observations	777	773	773	735	773

Notes: ***significant at 1% level; **significant at 5% level; *significant at 10%; corrected standard errors in parentheses.

Table B6. Determinants of expectations of trust, $E(\hat{S})$

	(1)	*(2)*	*(3)*
Pre-communication	33.649***	33.708***	37.726***
	(2.734)	(2.775)	(3.245)
Post-communication	24.233***	24.727***	27.629***
	(2.636)	(2.659)	(3.002)
Punishment	23.083***	23.542***	27.721***
	(2.719)	(2.753)	(3.313)
Repetition	22.867***	23.549***	26.979***
	(2.941)	(2.968)	(3.539)
Version (trust first)	1.462	2.291	2.080
	(3.992)	(3.922)	(3.876)
Female	–6.779	–9.285**	–7.251*
	(4.190)	(4.106)	(4.125)
White	3.081	3.231	2.028
	(4.231)	(4.208)	(4.189)
American	–2.801	0.249	–2.049
	(4.204)	(4.391)	(4.221)
Social orientation DG		0.231***	0.242**
		(0.058)	(0.119)
DG × pre-communication			–0.332***
			(0.119)
DG × post-communication			–0.262**
			(0.116)
DG × punishment			–0.378***
			(0.112)
DG × repetition			–0.317**
			(0.124)
Constant	43.261***	36.682***	41.614***
	(4.988)	(5.523)	(5.041)
R^2	0.102	0.104	0.108
N of clusters	194	192	192
Observations	932	922	922

Notes: ***significant at 1% level; **significant at 5% level; *significant at 10%; corrected standard errors in parentheses.

Table B7. Determinants of willingness to be trustworthy, $\hat{Y}$

	(1)	*(2)*	*(3)*	*(4)*	*(5)* *IV*
Pre-communication	49.861***	49.810***	56.222***	–0.926	
	(4.507)	(4.546)	(5.376)	(3.339)	
Post-communication	32.849***	33.135***	36.334***	–5.046	
	(4.725)	(4.787)	(5.606)	(3.264)	
Punishment	27.954***	28.058***	35.609***	–6.058	
	(4.646)	(4.697)	(5.848)	(4.073)	
Repetition	33.462***	33.844***	39.571***	–0.860	
	(5.303)	(5.363)	(6.495)	(3.910)	
Version (trust first)	–3.419	–2.071	–2.080	–4.509	–4.411
	(7.616)	(7.608)	(7.626)	(3.730)	(3.735)
Female	–23.655 ***	–24.633***	–24.639***	–11.621***	–12.263***
	(7.872)	(8.024)	(8.038)	(4.005)	(4.036)
White	12.693*	10.834	10.748	6.296	6.560
	(7.894)	(7.953)	(7.971)	(4.491)	(4.393)
American	–6.757	–5.996	–5.957	–4.249	–4.361
	(8.703)	(8.885)	(8.905)	(4.637)	(4.681)
Social orientation		–0.057	0.308	0.006	0.011
DG		(0.204)	(0.226)	(0.104)	(0.114)
DG × pre-communication			–0.518***	–0.043	
			(0.189)	(0.074)	
DG × post-communication			–0.246	0.129	
			(0.195)	(0.151)	
DG × punishment			–0.618***	–0.056	
			(0.176)	(0.100)	
DG × repetition			–0.459**	0.005	
			(0.205)	(0.129)	
Expectations of amounts sent				1.610***	1.532***
				(0.074)	(0.063)
Constant	67.389***	68.913***	64.357***	–5.266	–2.728
	(10.736)	(11.484)	(11.869)	(6.537)	(7.767)
R^2	0.0947	0.0953	0.0994	0.7481	0.745
N of clusters	194	192	192	192	192
Observations	916	906	906	902	902

Notes: *** significant at 1% level; ** significant at 5% level; * significant at 10%; corrected standard errors in parentheses.

IRIS BOHNET is a professor of public policy and the faculty chair of the Women and Public Policy Program at the Kennedy School of Government, Harvard University. A behavioral economist, her research focuses on trust and institutional design. She runs her experimental studies in different parts of the world, examining the causes and effects of trust from a cross-cultural and comparative institutional perspective. Bohnet has recently published in the *American Economic Review*, the *American Political Science Review*, *Experimental Economics* and the *Journal of Economic Behavior and Organization*, among others. She teaches decision analysis and negotiation at the Kennedy School.

ADDRESS: Kennedy School of Government, Harvard University, 79 JFK Street, Cambridge, MA 02138, USA [email: Iris_Bohnet@Harvard.edu].

YAEL BAYTELMAN is an associate researcher and professor of economics at the Economics Department, University of Chile. An applied micro-economist, her current research focuses on household consumption patterns. More generally, she is interested in social policy, income distribution and environmental and natural resource economics and policy. She started working on the current project with Iris Bohnet while a student at the Kennedy School of Government. She has published for UNDP on basic services social spending in Chile (1998), and trade and biodiversity for a local journal of the University of Chile (2003). Currently, she teaches Social Investment Appraisal.

ADDRESS: Departamento de Economía, Universidad de Chile, Diagonal Paraguay 257, Santiago, Chile [email: ybaytel@econ.uchile.cl].