Politics, Philosophy & Economics
000(00) 1–18
© The Author(s) 2011
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1470594X10387260
ppe.sagepub.com
SAGE

Trustworthiness is a social norm, but trusting is not

Cristina Bicchieri
University of Pennsylvania, USA

Erte Xiao
Carnegie Mellon University, USA

Ryan Muldoon
University of Western Ontario, USA

Abstract

Previous literature has demonstrated the important role that trust plays in developing and maintaining well-functioning societies. However, if we are to learn how to increase levels of trust in society, we must first understand why people choose to trust others. One potential answer to this is that people view trust as normative: there is a social norm for trusting that imposes punishment for noncompliance. To test this, we report data from a survey with salient rewards to elicit people's attitudes regarding the punishment of distrusting behavior in a trust game. Our results show that people do not behave as though trust is a norm. Our participants expected that most people would not punish untrusting investors, regardless of whether the potential trustee was a stranger or a friend. In contrast, our participants behaved as though being trustworthy is a norm. Most participants believed that most people would punish someone who failed to reciprocate a stranger's or a friend's trust. We conclude that, while we were able to reproduce previous results establishing that there is a norm of reciprocity, we found no evidence for a corresponding norm of trust, even among friends.

Keywords

trust, reciprocity, social norms, punishment

Corresponding author:
Ryan Muldoon, Stevenson Hall, London, ON N6A 5B8, Canada
Email: rmuldoo@uwo.ca

> It is impossible to go through life without trust: that is to be imprisoned in the worst cell of all, oneself.
>
> Graham Greene

Introduction

The purpose of this article is to distinguish between trusting as a behavior that is norm driven versus trusting as a behavior that is predicated on the anticipation of profit through reciprocation. This distinction is particularly important since trust and trustworthiness are important elements in personal, social, and economic exchanges. Without trust neither markets nor social relations could function and thrive, as nobody would willingly give something of value, be it money, goods, time, or sensitive information, if not for the expectation that the exchange will bear some fruit and will not leave the giver poorer or otherwise injured. The placement of trust allows actions that would not otherwise be possible and, depending upon the performance of the trustee, the truster may be better off or worse off than if he had not trusted. Trusting someone means making a choice that can either benefit both the truster and the trustee or benefit the trustee to a greater extent, but at the truster's expense.

According to Hardin's view of trust as encapsulated self-interest (2006), if A trusts B, she must have good reasons to do so. In other words, A trusts B to do x because it is in B's interest to fulfill A's trust. Being trustworthy in this context means not having an incentive to exploit A's trust. According to this view, when such incentives are absent, trusting makes no sense and is patently irrational. This view of trust draws a wedge between personal and *generalized* trust,[1] since someone who is embedded in a thick network of trusting relationships and experiences the trustworthiness of people around them may not have the propensity to extend trust to strangers. Trust as encapsulated self-interest may work in close, repeated interactions, or in any case in which one's actions are observable by others who may become partners in future interactions. But what about anonymous encounters or any situation in which the link between truster and trustee is not close, there is no transparency, and the possibility of sanctions is remote or just unfeasible?

An answer favored by some (Putnam, 1993) is that there is continuity between personal and generalized trust; those who are embedded in a thick network of trusting relationships and experience the trustworthiness of people around them will have the propensity to extend trust even to strangers. In this light, it would be expected that a person who is accustomed to particularistic trust would be disposed to generalize and invest resources even when faced with anonymous partners. The breeding ground of trust, cooperative habits, and solidarity is always the small group, and the move to generalized trust is something that happens almost by default, as a habit one does not shed just because the situation is unusual or different. There are, however, many examples of discontinuity between personal and generalized trust, and examples abound of societies in which individuals display high levels of trust or reciprocation among family members and small networks, but show complete mistrust of strangers, institutions, and other beneficiaries of generalized trust. There is even evidence that in countries with widespread corruption impersonal trust is a scarce good, whereas personal trust may flourish (Ensminger, 2001).

Yamagishi (2001) made an important distinction between trust and assurance that captures the above discontinuity. People within committed relations or stable groups feel safe with insiders because formal and informal sanctions (including ostracism) against a betrayer are strong enough. Assurance is precisely an expectation of trustworthiness of others based on an assessment of their interests and incentives. Assurance does not generalize to interactions or situations that are not 'guaranteed' by existing incentive structures. Trust, on the contrary, is meaningful only in situations characterized by a high level of social uncertainty, in which there are incentives to act dishonestly and the consequences of being the target of dishonesty are costly. Trust, in Yamagishi's view, is independent of an assessment of trustworthiness and is, rather, a generalized expectation about human benevolence.

Another way to detach trust from trustworthiness is to treat trust as a heuristic (Messick and Kramer, 2001). The idea is that we have developed simple rules to deal with specific classes of situations, and these rules in general produce satisfactory results. Thus, a default rule of trusting may be almost automatically applied in many situations in which a less automatic decision would require a significant expenditure of time and resources to gather information on one's partner, if that were at all possible. In this view, trusting occurs irrespective of one's expectations of others' trustworthiness, and thus is uncoupled from self-interest, and more generally from a consequentialist assessment of its potential outcomes. In a similar vein, Elster says that 'trust and solidarity are genuine phenomena that cannot be dissolved into ultra-subtle forms of self-interest' (1979: 146). Trust, in this case, may mean having a personal disposition to follow a social or even a moral norm, without a thought to the potentially negative effects that one's trusting actions may bring about.

We agree with Yamagishi in linking trust with unavoidable social uncertainty. That is, an agent faces social uncertainty when she believes her interaction partner has an incentive to act in a way that imposes costs (or harm) on her and she does not have enough information to predict if the partner will in fact act in such a way (Yamagishi and Yamagishi, 1994). In other words, when we trust, we know that the trustee's choice is unconstrained by such mechanisms as formal contracts, verbal commitments that can affect reputation, and explicit or implicit promises of future rewards or punishments. If any such mechanism were in place, then we would have good reasons to expect trustworthiness, but in this case it would make little sense to talk of 'trust'. Yet, if people are willing to trust even when they know the potential trustee has an incentive to behave opportunistically, are they behaving rationally?

Though we accept the distinction between assurance (or encapsulated self-interest) and trust, we want to claim that trusting can be rational, insofar as trusting acts as a signal, whose intended effect is to focus the recipient on a reciprocity norm. If such a norm exists and is shared, then it is rational to trust insofar as one believes that in so doing one will trigger reciprocation even when the material incentives to reciprocate are absent. Thus, Hardin's view that trusting is rational is vindicated only insofar as the truster's expectation of reciprocity plays a role in the decision to trust; however, this expectation is not necessarily generated by previous experience or interaction with the trustee or by assessing the trustee's self-interest. If, indeed, a reciprocity norm exists and is commonly shared, then it makes sense for the truster to try to focus the trustee on it, in the expectation that he will reciprocate and thus benefit the truster. Though trust may not be a social norm or a default rule, the existence of a strong reciprocity norm could be enough to support generalized trust.

Generalized trust is exemplified by a game that is meant to study behavioral trust, that is, the willingness to bet that another party will reciprocate (at a cost) a risky move. In most experiments with trust games, trusters invest around 50 percent of their money, contrary to game-theoretic predictions. Trustees on average repay close to the original investment. This behavior is difficult to explain within the usual game-theoretic models, as they all make the auxiliary hypothesis that agents are selfish and only care about their material payoffs. Reciprocal altruism has been advanced as a possible alternative explanation. According to this hypothesis, people will send money to the extent that they believe doing so will elicit reciprocity on the part of the other person. While reciprocal altruism would account for the fact that experimental participants frequently send and send back positive amounts of money, it also implies that that there should be a correlation between the proportion of the investor's endowment that is sent and the return rate – but there is none. Moreover, reciprocal altruism would predict no sending (or sending back) in instances in which the person you send to (or received from) is not the same as the person deciding whether (or who decided) to send money. Yet, in treatments in which this was the case, there was some sending (and sending back) although willingness to do either clearly declines when the tie between senders and receivers is indirect.

These results have led some behavioral scientists to claim that trust is norm driven (Dawes, 1991; Orbell and Dawes, 1991). This means that either (1) the act of trusting is an almost automatic response to a situation that, though different from the typical ones in which trust is normally bestowed, is perceived as similar enough to induce trusting behavior or (2) in a more conscious way, people believe that they are expected to trust, and that not trusting is form of behavior that is reproachable. Be that as it may, if trusting were a social norm, then we would expect a general agreement that lack of trust, unless justified by the situation, is blameworthy.

As we said at the outset, the goal of this article is precisely to distinguish between trusting as a behavior that is norm driven versus trusting as grounded on the anticipation of profit through reciprocation. If trusting were a social norm, then we would expect a general agreement that lack of trust would be punished. That is, individuals would expect non-trusting behavior to be penalized, even when they themselves would not be inclined to punish a transgressor. This is because a social norm, as opposed to a personal value, does not require one's allegiance; for a norm to exist, there must be a collective belief that the behavior dictated by the norm is widespread, as well as a shared belief that one is expected to engage in such behavior when appropriate and that transgressions might be punished. This means one may follow a norm of trust without a personal value or disposition to be trusting, simply because one expects others to follow it and one believes others think one ought to follow it. A social norm only requires a conditional preference for following it, a preference grounded on the right kind of expectations, not an unwavering commitment. It is therefore important, in order to determine whether a norm exists and applies to a particular situation, to elicit individuals' expectations about what *others* would do or expect one to do in that situation (Bicchieri, 2006). *Empirical expectations* about what most people similarly situated will or would do are important, since any norm (once it is in place) has a coordination function, and believing that none or very few follow it would deprive the norm of its power. Yet empirical expectations alone are not sufficient to induce compliance. Typically, pro-social norms do not reflect the immediate self-interest of their followers, and thus we also need a *normative expectation* that others think we should conform to the norm, and are prepared to punish us if we do not. Since compliance with a social norm depends upon the individual's expectations, even a person who would normally obey a trust norm may shirk it in anonymous encounters, where the weight of normative expectations is greatly diluted. Just observing behavior in such situations does not allow us to conclude that a norm does or does not exist, but there are other ways to explore this issue.

Asking people whether a specific behavior elicits condemnation or punishment is a better way to determine whether that behavior is dictated by a norm. Yet we cannot establish that there is a norm of trust only by asking people whether *they* would punish non-trusting behavior. Depending upon their personal values, some would punish no matter what and others would not. Some may feel a deep personal allegiance to a trust norm, whereas others may just abide by it in the right circumstances, but evade it whenever possible, and thus look upon transgressors with great indulgence. If we instead ask people what they expect others to do, then we have a clearer picture of what sort of behavior is socially required, provided individuals' expectations are in agreement. Only when there is such a consensus are we justified in claiming that a norm exists.

Note that a norm of trust may be contingent upon the relationship existing between the parties. It is entirely possible that, whereas there is no norm telling us to trust strangers, there exists a strong norm about trusting friends and family. In this case, lack of generalized trust may not be considered negative, but failing to trust a friend may elicit punishment. We thus designed an experiment aimed at eliciting participants' expectations about which behaviors, in the context of a trust game, bring about punishment, and as we shall see, lack of trust (either of a friend or of a stranger) is not one of them, but lack of reciprocity is.

Experiment

Experiment design

To determine the normative status of trust, we designed a survey with salient rewards to elicit participants' attitudes regarding the punishment of untrusting behavior. We began by supposing that if trust is a social norm, then participants should expect that untrusting behavior would be punished, following the account of social norms developed in Bicchieri (2006). Thus, our experimental aim was to elicit individuals' expectations about punishment so as to inform our understanding of whether trust is a norm.

To elicit the relevant expectations, we first described to the participants a previously conducted standard trust game (Berg et al., 1995). The standard trust game creates a situation in which one player must decide whether to trust another, who must then decide whether to honor or abuse this trust. Specifically, in the trust game, both investor and trustee receive an endowment of, for example, US$10. The investor transfers some, all, or none of his endowment to the trustee and the experimenter triples any amount sent. After observing the tripled amount, the trustee transfers back some, all, or none of the tripled amount to the investor. The investor earns his endowment of US$10, minus the transfer amount and plus any amount transferred back. The trustee earns the endowment of US$10, plus the tripled transfer amount and less any amount transferred back.

Trust is interpreted here as the willingness to bet that another player will reciprocate a risky move at a cost to themselves. Thus, a zero transfer amount suggests the investor does not trust the trustee at all and, similarly, a zero return amount suggests the trustee is not trustworthy.[2] Though economic theory predicts that in a one-shot, anonymous trust game there will be no trust and reciprocation, experimental data tell otherwise. In most experiments, participants do trust by investing around half of their endowment, but trustees' repayments are usually equal to or less than the original investment (Camerer, 2003). An interesting question is thus why so many individuals, when in the role of truster or investor, tend to 'trust' their counterpart when in fact it appears that trusting is costly. In particular, we test whether trust is a norm and thus people trust to obey the norm.

To test this, we described to the participants instances of the standard trust game in which the investor did not transfer any money and also instances in which the investor transferred some positive amount of money to the trustee. In these latter instances, we also described cases in which trustees did not return any money as well as cases in which the trustees returned some positive amount of money to the investor. For each scenario considered, we asked the participants two questions: whether they would impose a fine on either the investor or the trustee and what their expectations were about what the other participants would choose to do.

As noted before, we wished to consider whether the relationship between the investor and the trustee matters to the normative status of trust: it may be the case that trust among friends has a different status than trust among strangers. To investigate this possibility, we conducted both a 'stranger' treatment and 'friend' treatment of the full experiment. The only difference between these two treatments is that in the former the participants were told that the investor and the trustee are strangers, whereas in the friend treatment participants were told that the investor and the trustee are friends. By supplying this additional context about the relationship of the two parties, we aimed to activate any context-sensitive norms that may not have been triggered without such specification.

In both treatments, participants were first familiarized with the trust game, and then were asked to make judgments about whether they wanted to punish one of the actors in seven different scenarios, three of which focused on an investor's decision while the other four focused on a trustee's decision. The participants did not have to pay anything to punish the actors in the trust game. The details of the seven scenarios are listed in Table 1.

Each subject answered three questions within each scenario. First, participants were asked what fine (called a 'payoff cut' in the instructions) they would like to impose on the decision-maker described in the scenario. Possible options were 0 percent, 10 percent, 30 percent, 50 percent, 70 percent, 90 percent, and 100 percent of the actor's earnings. It was made clear to the subject that the fine's amount would not go either to the

Table 1. Scenarios to consider when making punishment decisions

Punishment target	Scenario to be considered	Denotation
Investor	The investor transferred US$0 to the trustee.	I(0)
	The investor transferred US$5 to the trustee.	I(1/2)
	The investor transferred US$10 to the trustee.	I(1)
Trustee	The investor transferred US$5 and the trustee returned US$0	R(0)
	The investor transferred US$5 and the trustee returned US$5	R(1/3)
	The investor transferred US$5 and the trustee returned US$10	R(2/3)
	The investor transferred US$5 and the trustee returned US$15	R(1)

subject herself or to the punished individual's counterpart. The money was taken away, but not redistributed.

The second question asked the subject to estimate how many participants in her session chose not to fine the decision-maker (that is, chose a fine of 0 percent). Participants were reminded that they would earn a point for giving a correct estimate. A third, related question was also asked: participants were asked to choose the punishment amount that most participants would choose in the subjects' session. As with the second question, the participants were made aware that correct choices would earn them one point. Each subject earned US$3 by completing all seven questions. At the end of the experiment, two questions were randomly selected from those for which participants could earn points. Participants earned US$3 per point on those two questions only (the Appendix provides details).

Experimental procedure

The participants were recruited at the University of Pennsylvania through the web-based 'Experiments @ Penn' recruitment system. They were given instructions for the trust experiment and were told that they would be asked to answer several questions regarding the participants' decisions in that experiment. Participants were given a short quiz to make sure that they understood the trust game. After each participant correctly completed the quiz, the experimenter handed out the punishment surveys. After all participants finished the surveys, two questions were randomly selected and participants were paid according to their answers to those questions. Each participant received a US$5 attendance bonus in addition to the money earned in the game and the survey (US$4 on average). Participants were in the laboratory for less than one hour.

Results

We obtained observations on 62 participants: 30 in the stranger treatment and 32 in the friend treatment. In each treatment k (with 'f' designating the friend treatment and 's' designating the stranger treatment), participant j indicated the amount of the fine to impose on the investor in each of the three scenarios *I(.)* or the trustees in each of the four scenarios *R(.)* (see Table 1). We obtained data on participant j's expectation regarding the percentage of participants who imposed no punishment on the decision-maker (investor or trustee) in each scenario. We denote this as $\mathrm{S}_{j,I(.),k}$ and $\mathrm{S}_{j,R(.),k}$, respectively. We also obtained the participants' expectations regarding the most frequently chosen fine amount, denoted as $P_{j,I(.),k}$ for the investor scenario and $P_{j,R(.),k}$ for the trustee scenario. For each scenario in each treatment, we calculated the average of each expectation across the participants: $\bar{S}_{I(.),k}$, $\bar{S}_{R(.),k}$, $\bar{P}_{I(.),k}$, and $\bar{P}_{R(.),k}$.

Previous research has shown that reciprocity is a social norm and third parties are often willing to punish violators (for example, Fehr and Fischbacher, 2004; Kurzban et al., 2006). Here, we compare punishment expectations between cases in which the investor is completely untrusting (that is, I(0)) and in which the trustee is completely untrustworthy (that is, R(0)). In addition, we compare results between the friend and stranger treatments to determine how a trust norm might vary according to the relationship between the investor and the trustee.

We begin with an analysis of the stranger treatment. We report participants' answers to questions regarding how many other participants they think will choose 'no punishment' in each of our several hypothetical investments and returns. If trust were a norm, then we would expect that, on average, individuals believe that a majority (more than 50 percent) will punish a completely untrusting investor (that is, $\bar{S}_{I(0),k} \geq 50$ percent). In fact, we find that, on average, participants believed that 60 percent of participants in their session would *not* impose any punishment on an investor who transferred zero (this is not statistically different than 50 percent, one-tail t-test $p = 0.944$). In contrast, only 24 percent (significantly less than 50 percent, one-tail t-test $p < 0.01$) of participants were expected not to impose a fine on a trustee who returned zero. The difference between these two mean expectations (60 percent and 24 percent) is significant (two-tail paired t-test, $p < 0.01$).

Figure 1(A) plots the distribution of $\bar{S}_{j,I(0),s}$ and $\bar{S}_{j,R(0),s}$. Significantly more participants expect at least 50 percent of participants to choose no punishment in the untrusting investor case than in the untrustworthy trustee case (60 percent versus 17 percent, two-tail paired t-test, $p < 0.01$). Moreover, 20 percent of participants (6 out of 30[3]) expected nobody to impose a fine on untrusting investors (that is, $\bar{S}_{j,I(0),s} = 100$), but only 3 percent of participants (1 out of 30) believed nobody would punish an untrustworthy trustee. On the other hand, while 30 percent (9 out of 30) believed at least some people would fine untrustworthy participants (that is, $\bar{S}_{j,R(0),s} = 0$), only 3 percent (1 out of 30) believed so for the untrusting investor. This evidence runs counter to the view that trust is a social norm.

Another way to measure whether trust is a norm is to ask how much punishment participants think most people would impose on an untrusting action. We next report data on the expectations that participants hold regarding the punishment level that most participants would choose. We plot the distribution of $\bar{P}_{j,I(0),s}$ or $\bar{P}_{j,R(0),s}$ in Figure 1(B). About 53 percent of participants expected zero punishment to be the most popular choice when an investor transfers zero. However, when the trustee returns nothing only about 17 percent of participants expected that most participants would choose no punishment. The average expected magnitude of the fine is also lower in the zero investment case than in the zero return case (21 percent versus 58 percent, respectively, two-tail paired t-test $p < 0.01$). This provides convergent evidence that choosing not to trust is not a norm violation.

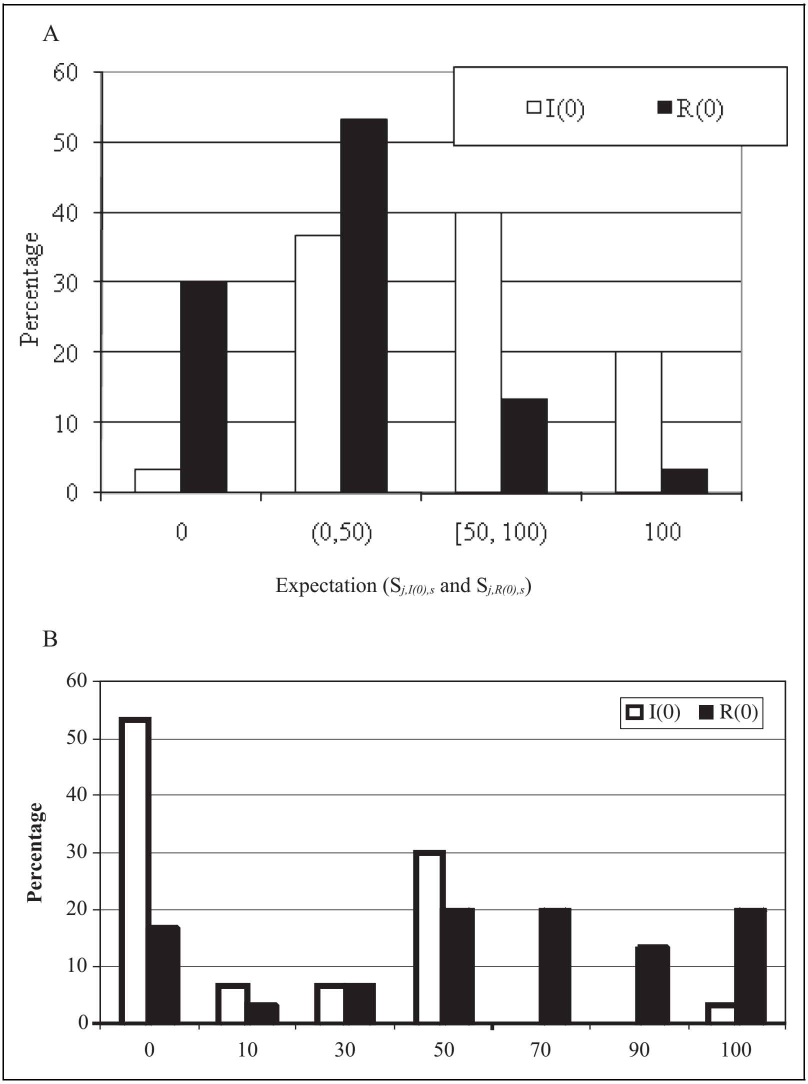

Figure 1. (A) Expectations of punishment in the stranger treatment: the distribution of expectations regarding the percentage of people who imposed no punishment on the investor when the investment amount is zero ($S_{j,I(0),s}$) or on the trustee when the amount returned is zero ($S_{j,R(0),s}$) (B) Expectations of punishment in the stranger treatment: the distribution of expectations regarding the most frequently chosen fine amount imposed on the investor when the investment amount is zero ($P_{j,I(0),s}$) and the amount imposed on the trustee when the amount returned is zero ($P_{j,R(0),s}$)

As we argued earlier, whether trust is a norm might vary with the relationship between the investor and trustee. In particular, a trust norm might generally exist among friends even if it does not exist among strangers. We were surprised to find that this does not seem to be the case. In particular, we conducted the same analysis in the friend treatment as in the stranger treatment discussed above. First, when the trustee and the investor are friends, on average, participants expected that 47 percent of the participants would not punish the investor at all if she transferred zero to her friend (this is not statistically significantly less than 50 percent (one-tail t-test $p = 0.342$) nor is it significantly different from 60 percent in the corresponding stranger treatment (two-tail t-test $p = 0.17$)). On the other hand, only 20 percent of participants were expected not to fine the trustee if she returned nothing to her friend (significantly less than 50 percent (one-tail t-test $p < 0.01$) and not significantly different from 24 percent in the corresponding stranger treatment (two-tail t-test $p = 0.50$)). The difference between these two expectations is significant (47 percent versus 20 percent, two-tail paired t-test, $p < 0.01$).

The distribution of $\bar{S}_{j,I(0),f}$ and $\bar{S}_{j,R(0),f}$ is plotted in Figure 2(A). Significantly more participants expect at least 50 percent of participants to choose no punishment in the untrusting investor case than in the case of an untrustworthy trustee (53 percent versus 16 percent, two-tail paired t-test, $p < 0.01$).

Figure 2(B) plots the distribution of $\bar{P}_{j,I(0),f}$ or $\bar{P}_{j,R(0),f}$ in the friend treatment. About 47 percent of participants expected that most participants would choose not to punish an untrusting investor at all. However, if the trustee returns nothing, only about 19 percent of participants expected most participants would choose not to punish the trustee. The average magnitude of the fine that most participants expected to be imposed is also lower in the zero investment case than that in the zero return case (30 percent versus 57 percent, two-tail pair t-test $p < 0.01$). Again, we do not find statistically significant differences between the stranger and friend treatments.

These results suggest that even when the trustee is the investor's friend, people do not expect that others would punish a decision not to trust the trustee. Thus, people do not seem to believe trust is a norm that people should obey, regardless of whether or not the investor and the trustee are friends.

We have discussed expectations regarding punishment decisions in cases in which either the investor shows no trust at all (sends nothing) or the trustee is completely untrustworthy (returns nothing). We next address how people expect punishment decisions to be made when the investor shows some degree of trust or the trustee is trustworthy to some degree. In our experiment, the investor in scenario I(1/2) signals some degree of trust, but not full trust, and in scenario I(1) the investor signals complete trust. In both cases, we found that, on average, about 90 percent of participants expected others not to punish the investor in the stranger treatment. About 80 percent of participants expected no punishment in the corresponding friend treatment.

Figure 3 plots $\bar{P}_{I(.),k}$ in all three investor cases I(0), I(1/2), and I(1) as well as in the four trustee cases R(0), R(1/3), R(2/3), and R(1) in both treatments. Note that there is no apparent difference between the stranger and the friend treatments.

If the investor transferred half of the endowment (scenario I(1/2)), the average expected fine amount chosen by most participants was close to zero. This was true regardless of whether the investor and trustee were friends or strangers. In particular, more than 90 percent of participants expected a zero fine to be chosen by most participants in both treatments. When the trustee returned the transfer amount (that is, R(1/3)), the expected fine was significantly lower than when the return amount was zero

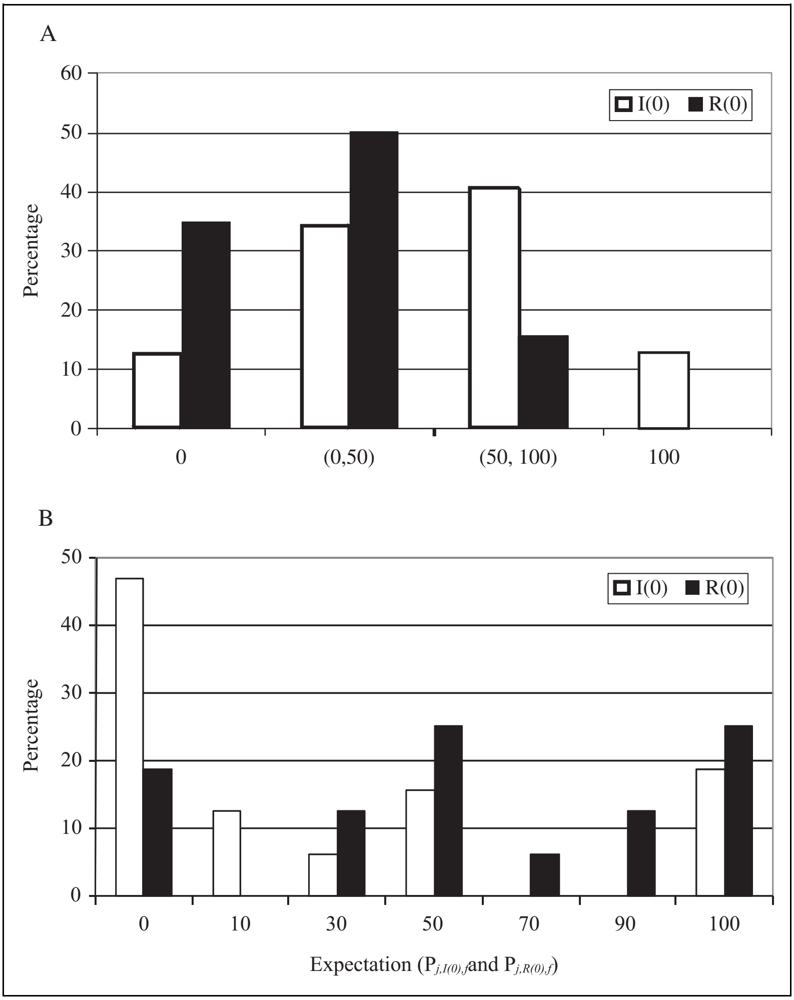

Figure 2. (A) Expectations of punishment in the friend treatment: the distribution of expectations regarding the percentage of participants who imposed no punishment on the investor when the investment amount is zero ($S_{j},_{I(0)},_{f}$) or on the trustee when the amount returned is zero ($S_{j},_{R(0)},_{f}$) (B) Expectations of punishment in the friend treatment: the distribution of expectations regarding the most frequently chosen fine amount imposed on the investor when the investment amount is zero ($P_{j},_{I(0)},_{f}$) and the amount imposed on the trustee when the amount returned is zero ($P_{j},_{R(0)},_{f}$)

(28 percent versus 58 percent, two-tail paired t-test $p < 0.01$ in the stranger treatment and 30 percent versus 57 percent, two-tail paired t-test $p < 0.01$ in the friend treatment). It is also interesting to notice that $\bar{P}_{R(1/3)},_{s}$ is not significantly different from $\bar{P}_{I(0)},_{s}$ (28 percent versus 21 percent, two-tail paired t-test $p = 0.29$) and that the same is true for $\bar{P}_{R(1/3)},_{f}$ versus $\bar{P}_{I(0)},_{f}$ in the friend treatment (both 30 percent). This suggests that returning the original investment amount, so that the investor is made whole, is not viewed as a norm

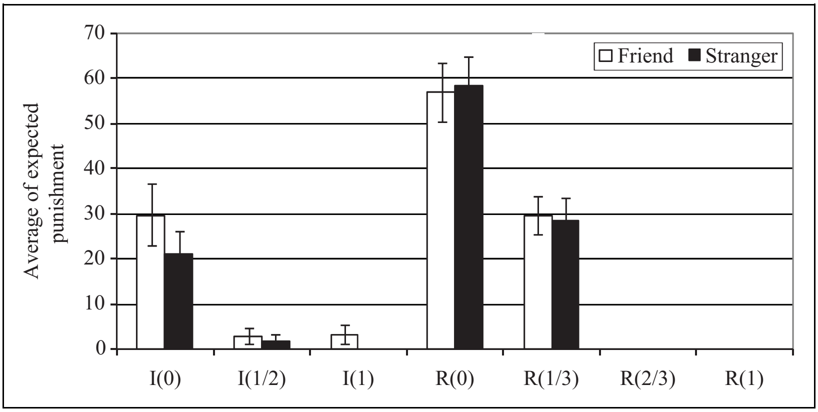

Figure 3. Average expected punishment amount imposed on the investor $\bar{P}_I(.)_{,k}$ and the trustee $\bar{P}_R(.)_{,k}$ in all seven scenarios in the stranger treatment and friend treatment

violation. Finally, in both treatments nobody expected any punishment to occur when the trustee returned the fair amount.

In sum, we obtained convergent evidence that people do not believe that to trust is a norm. Our participants expected that most people would not punish untrusting investors, regardless of whether the potential trustee was a friend or stranger. On the other hand, our participants behaved as though behaving in a trustworthy manner is a social norm: most of the participants believed that most people would punish someone who failed to reciprocate a stranger's or a friend's trust.

Discussion

Our results provide no evidence about the existence of a norm of trust. Interestingly, our data suggest that there is no difference in people's normative beliefs regarding trusting friends and trusting strangers. Often norms are contextual, in that they only cover a specific class of situations. So, for example, it may be the case that there is a general obligation to trust friends, but not strangers. What we observed, however, is that, even in close relationships, trusting does not seem to be normatively required. In contrast, punishment expectations show that there is a norm of reciprocity. This is not surprising. In every society such norms define certain actions and obligations as repayment for benefits received.

Trust is grounded upon reciprocity norms. Their very existence provides grounds for the expectation of being reciprocated. We expect people to help those who have helped them, and therefore we expect those whom we trust to have an obligation to honor our trust. This is the reason why we have argued that trusting can be rational, insofar as the truster expects, by her action, to focus the trustee on a reciprocity norm and thus trigger an adequate response. Nobody would trust without expecting to be at least made whole, but this expectation is a general one, not one specific to a particular party one knows has an interest in reciprocating.

Though our experiment only aimed to assess whether there is a trust norm, our results suggest that it is the presence of a norm of reciprocity that elicits trusting behavior in impersonal contexts. Interestingly, the fact that trustees usually return an amount equal to or less than the original investment may also be explained by a theory of norm compliance in which compliance is conditional upon having the right sort of expectations (Bicchieri, 2006). In the anonymous environment, there is no risk of being punished, and thus the pull of the norm, though present, is less strong.

Appendix: Instructions of the stranger treatment

Thank you for coming! You've earned $5 for showing up on time, and the instructions explain how you can make decisions and earn more money.

In today's session, every participant will be given a questionnaire. You will be asked to make decisions in several questions regarding participants' decisions in another experiment (Experiment A). Your payoff depends on your decisions in the questionnaire. To answer the questionnaire, you need to first understand Experiment A. The next page describes the experiment. Please read it carefully.

Description of Experiment A

In this experiment, two participants are paired up. One plays as Actor 1 and the other plays as Actor 2.

Actor 1 and Actor 2 are randomly and anonymously paired up. They will never be informed of each other's identity.

At the beginning of the experiment *both actors* receive an initial endowment of $10.

First, Actor 1, can transfer, from his/her endowment, any amount from $0 to $10 to Actor 2. The experimenters will *triple* this transferred amount, so that Actor 2 receives three times the number of dollars Actor 1 transferred.

Then, after Actor 1's decision, Actor 2 can transfer back to Actor 1 any amount of the tripled number of dollar bills he/she received.

Final payoffs. *Actor 1* receives: $10 – transfer to Actor 2 + back-transfer from Actor 2.

Actor 2 receives: $10 + 3 × transfer from Actor 1 – back-transfer to Actor 1.

To ensure you understand Experiment A, please complete the following exercise:

1. At the beginning of the experiment, ___ receive an initial endowment of $10.
   a) Actor 1 b) Actor 2 c) Actor 1 and Actor 2 d) No one
2. If Actor 1 transfers $5 and Actor 2 transfers $10 back to Actor 1, then
   Actor 1's final payoff = $___
   Actor 2's final payoff = $___
3. In Question 2, if Actor 2 transfers $1 back to Actor 1, then
   Actor 1's final payoff = $___
   Actor 2's final payoff = $___

Note: Both Actor 1 and Actor 2 receive an initial endowment of $10.

Final payoffs. Actor 1 receives: $10 – transfer to Actor 2 + back-transfer from Actor 2.
Actor 2 receives: $10 + 3 × transfer from Actor 1 – back-transfer to Actor 1.

Questionnaire

Please read the following questions carefully, you will earn $3 by finishing all the questions. You will earn points from some of the questions. At the end of the experiment, two questions will be randomly selected from those for which you can earn points. You will earn $3 for each point you earned.

In order to answer the following questions, you need to know there are ___ participants in total in today's session.

ID: In the following questions, you will decide whether to impose a payoff cut to *Actor 1*'s final payoff in each scenario. The payoff cut amount is the number of *cents* you would deduct from *each dollar* of Actor 1's earnings. The payoff cut amount does *NOT* go to either Actor 2 or you.

In order to answer the following questions, you need to know there are ___ participants in total in today's session.

Scenario: Actor 1 transferred $0 to Actor 2.

- I would choose ___.
  a) no payoff cut to Actor 1
  b) to deduct 10 cents from each dollar of Actor 1's earnings
  c) to deduct 30 cents from each dollar of Actor 1's earnings
  d) to deduct 50 cents from each dollar of Actor 1's earnings
  e) to deduct 70 cents from each dollar of Actor 1's earnings
  f) to deduct 90 cents from each dollar of Actor 1's earnings
  g) to deduct all the earnings from Actor 1

Please briefly explain your decision here.

- How many participants in today's session do you think chose 'a) no payoff cut to Actor 1'?
  (If your answer is right, you will get one point.)
- What is the option that you think most participants chose today?
  (If your answer is right, you will get one point.)

Scenario: Actor 1 transferred $5 to Actor 2.

- I would choose ___.
  a) no payoff cut to Actor 1
  b) to deduct 10 cents from each dollar of Actor 1's earnings
  c) to deduct 30 cents from each dollar of Actor 1's earnings
  d) to deduct 50 cents from each dollar of Actor 1's earnings
  e) to deduct 70 cents from each dollar of Actor 1's earnings
  f) to deduct 90 cents from each dollar of Actor 1's earnings
  g) to deduct all the earnings from Actor 1

Please briefly explain your decision here.

- How many participants in today's session do you think chose 'a) no payoff cut to Actor 1'?
  - (If your answer is right, you will get one point.)
- What is the option that you think most participants chose today?
  - (If your answer is right, you will get one point.)

Scenario: Actor 1 transferred $10 to Actor 2.

- I would choose ___.
  - a) no payoff cut to Actor 1
  - b) to deduct 10 cents from each dollar of Actor 1's earnings
  - c) to deduct 30 cents from each dollar of Actor 1's earnings
  - d) to deduct 50 cents from each dollar of Actor 1's earnings
  - e) to deduct 70 cents from each dollar of Actor 1's earnings
  - f) to deduct 90 cents from each dollar of Actor 1's earnings
  - g) to deduct all the earnings from Actor 1

Please briefly explain your decision here.

- How many participants in today's session do you think chose 'a) no payoff cut to Actor 1'?
  (If your answer is right, you will get one point.)
- What is the option that you think most participants chose today?
  (If your answer is right, you will get one point.)

ID: In the following questions, you will decide whether to impose a payoff cut to *Actor 2*'s final payoff in each scenario. The payoff cut amount is the number of *cents* you would deduct from *each dollar* of Actor 2's earnings. The payoff cut amount does *NOT* go to either Actor 1 or you.
In order to answer the following questions, you need to know there are ___ participants in total in today's session.
Scenario: Actor 1 transferred $5 and so Actor 2 received $15. Then Actor 2 transferred back $0. Therefore, at the end of the experiment, Actor 1 received $5 and Actor 2 received $25.

- I would choose ___.
  - a) no payoff cut to Actor 2
  - b) to deduct 10 cents from each dollar of Actor 2's earnings
  - c) to deduct 30 cents from each dollar of Actor 2's earnings
  - d) to deduct 50 cents from each dollar of Actor 2's earnings
  - e) to deduct 70 cents from each dollar of Actor 2's earnings
  - f) to deduct 90 cents from each dollar of Actor 2's earnings
  - g) to deduct all the earnings from Actor 2

Please briefly explain your decision here.

- How many participants in today's session do you think chose 'a) no payoff cut to Actor 2'?
  (If your answer is right, you will get one point.)
- What is the option that you think most participants chose today?
  (If your answer is right, you will get one point.)

Scenario: Actor 1 transferred $5 and so Actor 2 received $15. Then Actor 2 transferred back $5. Therefore, at the end of the experiment, Actor 1 received $10 and Actor 2 received $20.

- I would choose ___.
  a) no payoff cut to Actor 2
  b) to deduct 10 cents from each dollar of Actor 2's earnings
  c) to deduct 30 cents from each dollar of Actor 2's earnings
  d) to deduct 50 cents from each dollar of Actor 2's earnings
  e) to deduct 70 cents from each dollar of Actor 2's earnings
  f) to deduct 90 cents from each dollar of Actor 2's earnings
  g) to deduct all the earnings from Actor 2

Please briefly explain your decision here.

- How many participants in today's session do you think chose 'a) no payoff cut to Actor 2'?
  (If your answer is right, you will get one point.)
- What is the option that you think most participants chose today?
  (If your answer is right, you will get one point.)

Scenario: Actor 1 transferred $5 and so Actor 2 received $15. Then Actor 2 transferred back $10. Therefore, at the end of the experiment, Actor 1 received $20 and Actor 2 received $10.

- I would choose ___.
  a) no payoff cut to Actor 2
  b) to deduct 10 cents from each dollar of Actor 2's earnings
  c) to deduct 30 cents from each dollar of Actor 2's earnings
  d) to deduct 50 cents from each dollar of Actor 2's earnings
  e) to deduct 70 cents from each dollar of Actor 2's earnings
  f) to deduct 90 cents from each dollar of Actor 2's earnings
  g) to deduct all the earnings from Actor 2

Please briefly explain your decision here.

- How many participants in today's session do you think chose 'a) no payoff cut to Actor 2'?
  (If your answer is right, you will get one point.)
- What is the option that you think most participants chose today?
  (If your answer is right, you will get one point.)

Scenario: Actor 1 transferred $5 and so Actor 2 received $15. Then Actor 2 transferred back $15. Therefore, at the end of the experiment, Actor 1 received $20 and Actor 2 received $10.

- I would choose ___.
  a) no payoff cut to Actor 2
  b) to deduct 10 cents from each dollar of Actor 2's earnings
  c) to deduct 30 cents from each dollar of Actor 2's earnings
  d) to deduct 50 cents from each dollar of Actor 2's earnings
  e) to deduct 70 cents from each dollar of Actor 2's earnings
  f) to deduct 90 cents from each dollar of Actor 2's earnings
  g) to deduct all the earnings from Actor 2

Please briefly explain your decision here.

- How many participants in today's session do you think chose 'a) no payoff cut to Actor 2'?
  (If your answer is right, you will get one point.)
- What is the option that you think most participants chose today?
  (If your answer is right, you will get one point.)

Notes

We wish to thank Alex Chavez and Gerry Mackie for comments.

[1] What is meant by 'generalized trust' is not just trust extended to random others, but also trust extended to impersonal actors such as social institutions, without grounding such trust in prior relationships or in the possibility of monitoring and sanctioning lack of trustworthiness.

[2] Cox (2004) designed a triadic game to sort out the motivations of investors and trustees in the trust game. He found that part of the reason that investors transfer a positive amount may be due to altruistic motivation and that trustees may also return a positive amount due to inequality aversion. Nevertheless, we may still assume that a zero transfer amount signals no trust and that a zero return signals no reciprocity. Our conclusions are thus drawn from the data for the cases of zero transfer and zero return.

[3] Of these six participants, two believed nobody would choose to impose zero punishment on the trustee when she returned zero.

References

Berg J, Dickhaut J and McCabe K (1995) Trust, reciprocity, and social history. *Games and Economic Behavior* 10: 122–42.

Bicchieri C (2006) *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

Camerer C (2003) *Behavioral Game Theory*. New York: Russell Sage Foundation.

Cox J (2004) How to identify trust and reciprocity. *Games and Economic Behavior* 46: 260–81.

Dawes R (1991) Social dilemmas, economic self-interest and evolutionary theory. In: Brown DR, Smith JEK (eds) *Recent Research in Psychology: Frontiers of Mathematical Psychology: Essays in Honor of Clyde Coombs*. New York: Springer-Verlag, 53–79.

Elster J (1979) *Ulysses and the Sirens: Studies in Rationality and Irrationality*. New York and Cambridge: Cambridge University Press.

Ensminger J (2001) Reputations, trust, and the principal-agent problem. In: Cook K (ed.) *Trust and Society*. New York: Russell Sage Foundation.

Fehr E and Fischbacher U (2004) Third-party punishment and social norms. *Evolution and Human Behavior* 25: 63–87.

Hardin R (2006) *Trust*. New York: Wiley.

Kurzban R, DeScioli P and O'Brien E (2006) Audience effects on moralistic punishment. *Evolution and Human Behavior* 28: 75–84.

Orbell J and Dawes R (1991) A 'cognitive miser' theory of cooperators' advantage. *American Political Science Review* 85(2): 515–28.

Putnam R (1993) *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.

Yamagishi T and Yamagishi M (1994) Trust and commitment in the United States and Japan. *Motivation and Emotion* 18: 129–66.

About the Authors

Cristina Bicchieri is the Carol and Michael Lowenstein Professor of Philosophy and Legal Studies at the University of Pennsylvania. She works on judgment and decision-making, with a special interest in decisions about fairness, trust, and cooperation, and how expectations affect behavior.

Erte Xiao is an Assistant Professor in the Department of Social and Decision Sciences at Carnegie Mellon University. She received her PhD in Economics at George Mason University and spent two years at the University of Pennsylvania as a post-doctoral fellow. Her published articles focus on understanding how incentives, social norms, and emotions affect individual decision-making in social and economic exchange environments.

Ryan Muldoon is a Post-doctoral Fellow at the Joseph L. Rotman Institute of Science and Values in the Philosophy Department at the University of Western Ontario. His research focuses on diversity in the social contract, norm dynamics, and the division of cognitive labor.