Fall 12-3-2018

Why Trust Out-groups? The Role of Punishment Under Uncertainty

Xiaofei Pan
Bryant University, xpan@bryant.edu

Daniel Houser
George Mason University, dhouser@gmu.edu

Why trust out-groups? The role of punishment under uncertainty

Xiaofei Pan[1] and Daniel Houser[2]

We conducted a hidden-effort trust game, in which we assigned subjects to one of two groups. The groups, which were formed through two different group formation processes, included a "social" group that required sharing and exchange among its members, and a "non-social" group that did not. Once assigned, subjects participated in the game with members from both groups, either with or without the opportunity to punish a trustee who may have defected on them. We found that for investors in the non-social group, the opportunity to punish a trustee worked to promote trust, but only when the trustee was a member of the other group. For the social group, the opportunity to punish had no effect on the investors' trust decisions, regardless of the trustee's group. We provide a theoretical framework to explain this asymmetric effect of punishment on trust. Our results suggest that groups with identities founded in sharing and exchange—a feature of globalized societies—may find it less necessary to engage in costly punishment. As a result, they may enjoy gains in economic efficiency.

Key words: group identity, trust, punishment, belief, uncertainty

JEL: C91, D6

[1] Xiaofei Pan is an Assistant Professor of economics at Bryant University, 1150 Douglas Pike, Smithfield, RI, xpan@bryant.edu; Daniel Houser is a Professor of economics at George Mason University, 4400 University Drive, Fairfax, VA, 22030. dhouser@gmu.edu.

[2] We thank Tim Cason, Gary Charness, Yan Chen, David Eil, Charlie Plott for their helpful comments. We also thank seminar participants at ESA, University of California at Santa Barbara, Central South University, Nanyang Technological University, National University of Singapore, and University of Pennsylvania.

1. Introduction

Across organizations, countries and cultures, individuals exhibit varying levels of trust and cooperation with those they regard as in-group or out-group members (Alesina Baqir and Easterly 1999, Bohnet et al. 2012, Buchan et al. 2009, Fukuyama 1995, Yamagishi and Yamagishi 1994). This can have important consequences, as the decision to trust or not trust others plays an important role in determining economic growth (Knack and Keefer 1997, Zak and Knack 2001, Bottazzi Da Rin and Hellmann 2011). In part, this decision can stem from a group's social norms, and the way these norms prescribe interactions with in- and out- groups[3]. For example, Fukuyama (1995) found that while a lineage-based family business in Southern China grew rapidly due to trust among group members, the family's lack of trust of the out-group (those outside their lineage) created barriers to further expansion. On the other hand, Buchan et al (2009) reported that people from more globalized countries behave cooperatively towards both in- and out- group members. An explanation is that people from the globalized countries recognize the benefits of sharing and exchange, and thus are more likely to draw broad group boundaries[4].

Group identity defines the norms to which a group adheres (Akerlof and Kranton 2005). Eckel and Grossman (2005) found that groups created with interaction and cooperation during the group formation process are likely to believe in the cooperativeness of others and to contribute more in public goods games (Eckel and Grossman 2005). Group identity formation has also been shown to influence social preferences (Chen and Li 2009) and to impact the ability to coordinate (Charness et al 2007). Nonetheless, the literature has not yet examined whether a particular group identity formation process can influence members' willingness to trust. Further, the literature is silent on the outcome of interactions between members of groups created under different processes. Specifically, little is known about how investors' beliefs about the likelihood of reciprocity differ according to the social nature of their group identity, or how investors' belief differences influence their interaction with in- or out-group members.

[3] Social identity theory was first developed by Tajfel and Turner (1979) and serves as a psychological basis for intergroup discrimination. Akerlof and Kranton (2000, 2002, 2005) were first to introduce social identity into economic analyses.

[4] Globalization strengthens cosmopolitan attitudes by weakening the relevance of ethnicity, locality and nationhood as sources of identification (Buchan et al. 2009).

Beliefs about the likelihood of reciprocity can also be influenced by punishment institutions. It is well-established that punishment can promote cooperation (e.g. Fehr and Gachter 2000; Houser et al. 2008[5]). A reason is that the ability to punish increases one's confidence in other's cooperativeness (or trustworthiness). This can be true even when interacting with counterparts from an out-group (Meyerson et al, 1996; Zucker, 1986).

A goal of this paper is to disentangle the impact of a group's identity from the presence of punishment opportunities on investors' trusting behaviors and their beliefs about in- or out-group trustees' likelihood of reciprocity. It is difficult to separate these effects using data from natural environments. The reason is that the presence of punishment may be jointly determined with a group's identity during the group's formation process. For this reason, we conducted our study using controlled laboratory experiments.

In our experiment, we randomly assigned participants to an environment with or without punishment opportunities. We also randomly assigned them to one of two different group formation processes: a "social" process that involved substantial sharing and exchange among group members, and a "non-social" practice that required little or no sharing and exchange.

We developed and tested the hypothesis that punishment may have different effects on groups formed in these two different ways. We hypothesized that members of a group formed under a more social process may be more inclined to draw broader group boundaries and thus treat out-group members similarly to in-group members, even absent protection from a punishment institution. In contrast, a group formed without a social process might have narrower group boundaries and thus be reluctant to trust out-group members without the protection of punishment. As a result, punishment opportunities may not increase a social group member's likelihood to trust an out-group member, but may increase a non-social member's willingness to trust a member of an out-group. We refer this hypothesis as the *"Asymmetric effect of punishment on trust."*

To test this hypothesis, our experiment included both a group formation stage and a trust game stage. In the group formation stage, we created two distinct group experiences through a puzzle game that varied in terms of the level of sharing and exchange needed for successful

[5] Several previous studies have reported detrimental effects of punishment on trust (see, e.g., Fehr and Rockenbach 2003; Houser et al, 2008).

completion (Pan and Houser 2013, Eckel and Grossman 2005). One group was formed by working on a task that, in relation to the other group's task, required significantly more cooperation (i.e., more sharing and exchanges among the group members as in globalized societies). We refer to the former group as *Social* and the latter as *Non-Social*.

In the trust game stage, we randomly assigned participants to play the role of either an investor or a trustee in a two-period hidden-action trust game (a variant of the game in Charness and Dufwenberg 2006). The game was similar to a standard trust game, in that an investor could provide resources to a trustee, who then had an opportunity to reciprocate; however, our game added the possibility for defection, either by chance or as a result of a trustee choosing not to reciprocate. Although the investor knew this, she was not told the trustee's decision.[6] The hidden-action game was played in environments where punishment was possible and where it was not. In the *Punishment* condition, if the investor received a defection outcome, she could then reduce a trustee's earnings at no cost to herself. We did not allow investors to punish if the trustee reciprocated (this eliminates antisocial punishment as reported, for example, by Herrmann et al 2008). Our experiment design is thus two-by-two, with two group identities (*Social, Non-social*) and two punishment conditions (*Punishment* and *No Punishment*).

We found clear evidence supporting the *Asymmetric Effect of Punishment* hypothesis. Specifically, we found that when punishment opportunities are available, a *Non-social* group investor is more willing to trust an out-group trustee than when punishment is absent. By contrast, *Social* investors trust out-group trustees no differently when punishment is available than when it is not. On the other hand, in neither group does punishment itself affect the investor's trust towards an in-group trustee. We provide a belief-based theoretical framework to explain these observations, and show that it is consistent with beliefs we elicited from our participants. Finally, apart from our main investigation into the effect of punishment on trust, we also investigate whether investors from different groups might adopt different punishment strategies after receiving a defection payoff. We found that the *Social* investors use punishment significantly less frequently than their *Non-social* counterparts.

Our paper is divided into six parts. Section 2 discusses the recent research related to group identity formation and use of punishment. Section 3 introduces the experiment design, and

[6] We use "she" to refer to an investor and "he" to refer to a trustee.

Section 4 describes our predictions. Section 5 presents our experimental results; Section 6 discusses and concludes.

2. Related literature on group identity and punishment

Buchan et al. (2009) found that individuals from globalized societies draw broader group boundaries than individuals from countries with less globalization. As a result, these individuals are more cooperative and trusting and contribute more to public goods games[7]. Previous studies have also shown that different group formation processes can potentially influence members' subsequent economic decisions. For identities created in the laboratory, Eckel and Grossman (2005) found that actions designed to enhance team identity, such as group problem solving, contributed to higher levels of team cooperation in public goods games. Charness, Rigotti and Rustichini (2007) found that only salient group membership affects individual behavior, while minimal group paradigm[8] alone has little effect on individual behavior. Other studies have found that different experiences may have spillover impact on later economic decisions. Peysakhovich and Rand (2016) found in a laboratory experiment that individuals who experienced more cooperative environments were more likely to be prosocial in a subsequent game.

Previous studies examining the effect of group identity suggest that group identify can be formed through common belief systems. Masella et al. (2014) showed that an agent's group identity influences their belief about principals' controlling decisions and consequently influences agents' transfer decisions (similar to reciprocity). We thus view our paper as a complement to theirs. Ockenfels and Werner (2014) focused on the role of beliefs in one's giving behavior. They found that knowing whether the recipient shares identity with the dictator is important, in addition to knowing whether the dictator believes the recipient knows this. We again view our paper acts as a complement to this work. We contribute to this literature by studying how group identity influences "trust" decisions through the channel of beliefs[9].

[7] Rand et al (2008) illustrated that changing conflicts would make one draw different group boundaries to define their alliances.

[8] Tajfel and Turner (1986) described a set of criteria required for a group classification to be minimal. The conditions are: a) Subjects are randomly assigned to non-overlapping groups on the basis of some trivial tasks; b) No social interaction (like face-to-face or online chat) takes place between the subjects; c) Group membership is anonymous; and d) The decision tasks requires no link between a chooser's self-interest and her choices. Creation of salient group membership often relaxed one or more of the above requirements.

[9] In a meta-analysis by Lane (2016), beliefs are also found to influence in-group favoritism decisions.

We followed the design of Eckel and Grossman (2005) by using their puzzle game and shuffling the puzzle pieces among the group members. In doing so, we aimed to create exchange and sharing (as occurs in globalized environments). By controlling the extent of sharing and exchange needed for puzzle completion, our design aimed to create different sharing and exchange norms among members of different groups, and potentially different group boundaries (Buchan et al 2009).

Previous research has also addressed punishment in trust games[10]. Some studies have found detrimental effects of sanctions, showing that they can reduce trustee reciprocity (Fehr and Rockenbach 2003; Falk and Kosfeld 2006, Houser et al. 2008). Very little research has studied the effect of punishment opportunities on investors' trust decisions. Some works support the idea that punishment has a positive impact on people's willingness to trust and cooperate. Yamagishi (1986) showed that people favor sanctioning systems when there is little trust in others' cooperation. Knowing that one can punish a defector results in confidence and increases one's willingness to trust. Some other papers report that sanctioning institutions may negatively affect trust. Mulder Dijk Cremer and Wilke (2006) found that participants who have experienced sanctioning systems can potentially be less trusting than those who have not.

Previous studies of the impact of identity on punishment have obtained mixed results. These studies often focus on third party punishment of a norm violator when the third party either does or does not share the victim's identity (Goette et al 2006, Meier et al 2012a, 2012b). Bernhard et al (2006) studied a variety of conditions under which the third party punisher, the norm violator and the victim carry the same or different identities. Results are also mixed regarding how second parties punish violators (Chen and Li 2009; Weng and Carlsson 2012).

3. Experiment Design

Our experiment builds upon previous research on group identity formation (Eckel and Grossman (2005) and the hidden action trust game (Charness and Dufwenberg 2006). We used a two-by-

[10] Much literature has focused on 3rd-party punishment behavior (Bernhard et al 2006; Geotte et al 2006). Chen and Li (2009) studied punishment behavior from the 2nd party and found that one is more likely to forgive an in-group's misbehavior.

two design: the two groups were *Social*, *Non-social*[11]; crossed with the two incentive conditions *Punishment* and *No Punishment*[12]. We obtained decisions from a total of 278 subjects. 136 were randomly assigned to the *Social* group, with 68 participating in the *No Punishment* condition and 68 participating in the *Punishment* condition. The remaining 142 subjects were randomly assigned into the *Non-social* group, with 70 participating in the *No Punishment* condition and 72 participating in the *Punishment* condition.

Our experiment paradigm consisted of two stages. The first stage was the group formation stage. Subjects assigned to the *Social* group had to solve a triangle puzzle, while subjects assigned to the *Non-social* group had to solve a square puzzle. We shuffled the puzzle pieces among the group members so that to finish his/her own puzzle, a group member had to find the right pieces from his/her other group members. In the *Social* condition, each group member received a sealed envelope containing three of the four unique pieces necessary to complete the puzzle, in addition to one duplicate piece (see Figure 1a). This made necessary significant cooperation and exchange of puzzle pieces among the group members. By contrast, those assigned to the *Non-social* group worked on a square puzzle that required less cooperation and less exchange of pieces among the group members. The reason is that each piece included a corner of the square, leaving it straightforward to solve the puzzle[13]. A group was not considered finished until all group members had completed their task. The group that finished first received an additional $2 for each group member. To avoid the effect of wealth differences on decisions in the trust game, subjects did not learn that they had won until they completed making their trust game decisions (see Puzzle instruction in Appendix A).

[11] In each session, half of the subjects were randomly assigned to the *Social* group and the other half to the *Non-social* group. There were either 8 or 10 subjects in each session.

[12] The results of the *No Punishment* condition were initially reported in Pan and Houser (2013). The current paper studies the role of punishment on willingness to trust, and whether *Social* and *Non-social* groups use and respond to punishment differently. Pan and Houser (2013) focuses exclusively on how the group formation process affects trust decisions.

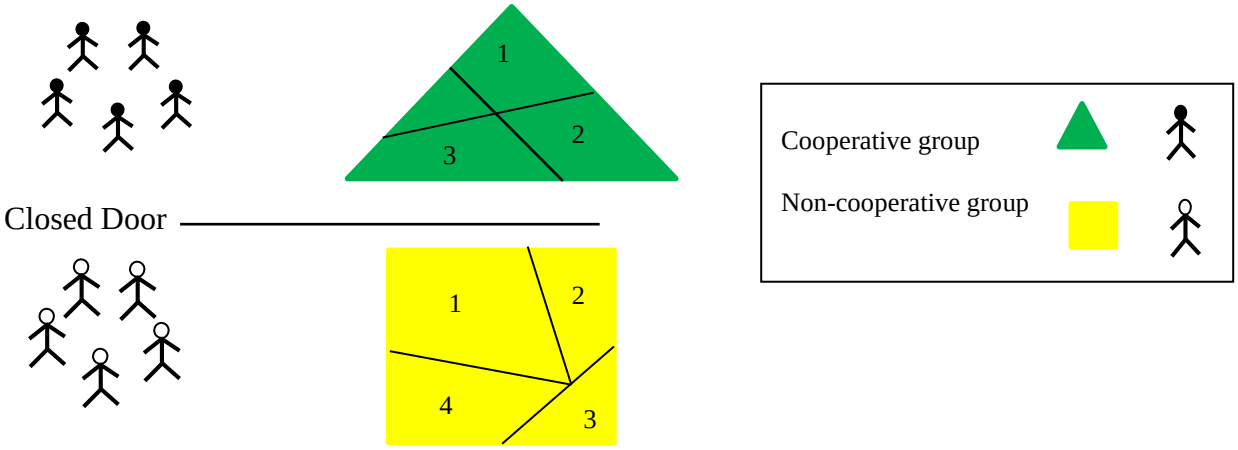
[13] Those who worked on the social task (Triangle puzzles) spent significantly more time than those who worked on the non-social task (square puzzles) (218 seconds vs. 102 seconds, $p < 0.05$). Further, we recorded the number of puzzle piece exchanges by comparing the specific pieces used in each member's finished puzzle to the specific pieces initially distributed to each person. People who spent more time finishing the task (also more likely to be in the *Social* group engaged in more exchange of puzzle pieces (mean = 72% new pieces), compared with those who exchanged fewer pieces (mean = 30%, $p < 0.01$).

After the group formation stage, participants played the trust game for two periods. Participants knew that they would be randomly matched with a different counterpart in each period. They also knew that one of the periods would be randomly chosen to determine the pair's earnings. The trust game was based on the hidden action trust game used by Charness and Dufwenberg (2006, Figure 2a). We used the strategy method[14] to elicit participants' decisions regarding in-group and out-group members.

At the beginning of the first period, each investor made two decisions about whether to trust (i.e., to choose IN or OUT). In one decision they were asked to assume they were matched with an in-group trustee, and in the other they were asked to assume they were matched with an out-group trustee. Similarly, trustees were asked to decide whether to reciprocate (i.e., choose to roll a die or not) if they were matched with an in- or an out- group investor. The game was a hidden effort trust game in that reciprocation by trustees was not observable to investors: there was a one-in-six chance that reciprocating would lead to the same outcome as defection.

Figure 1. Group formation stage

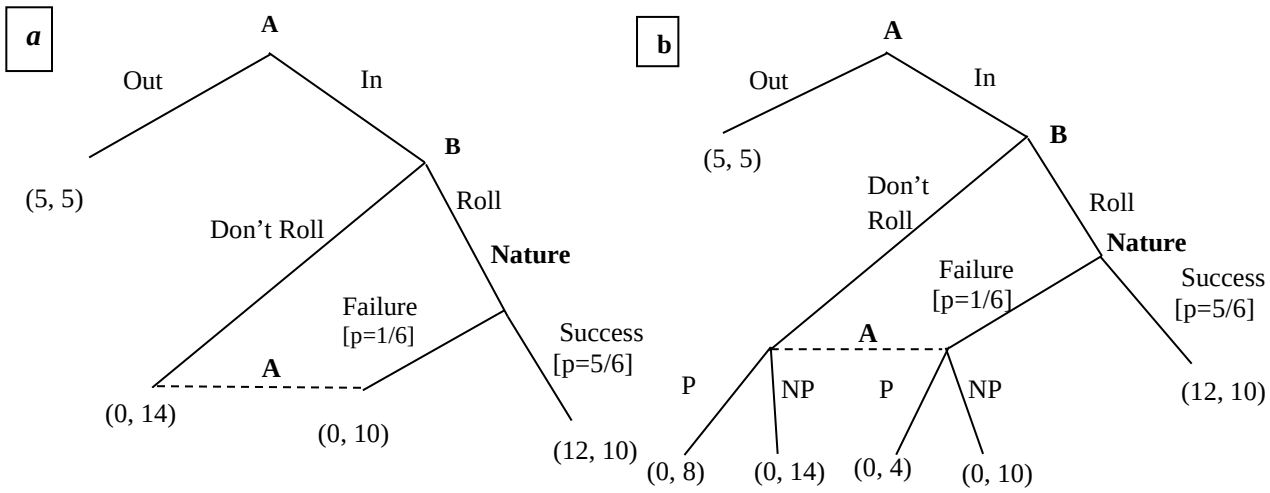
Each triangle puzzle was cut with only one right angle (as seen as piece #1). Each square puzzle included a right angle on each piece, making it relatively easier to solve.



[14] Strategy method allowed us to collect more data and conduct within subject comparison. The main disadvantage of the strategy method is that it could potentially diminish the effect of emotions. However, as our focus was to compare between the groups formed under social and non-social environments, such diminished effect, if any, would have been identical across the two conditions.

Figure 2: **Hidden action trust game.**

Figure 2*a*. Hidden action trust game absent punishment. Figure 2*b*. Hidden action trust game with punishment. *NP* represents Not to punish, *P* represents punish. In both Figure *a* and *b*, nature decides, if B ever rolls, it is a success or a failure. A represents an investor, and B represents a trustee.



In Period Two, the investors and trustees were matched with new counterparts. They were asked to make decisions in Period 2 conditional on all possible Period 1 outcomes. For example, if an investor chose IN for both an in-group trustee and an out-group trustee in Period 1, then this would result in four possible outcomes in period 2, as described below.

Period 2: Four outcome scenarios for an investor who chose to trust both an in-group and out-group member in Period 1[15]

1. If you received \$0 from your in-group[16] trustee
2. If you received \$12 from your in-group trustee
3. If you received \$0 from your out-group trustee
4. If you received \$12 from your out-group trustee

[15] On the other hand, those who only chose IN for in-group and OUT for out-group trustee could face only three possible scenarios: the first two would be identical to the table above, while the third (also the last) would be: "If you received \$5 from your out-group trustee." Similarly, those who chose OUT for both in- and out-group trustee would only face two scenarios.

[16] The word "in-group" is replaced by the respective group the participant belongs to. For instance, an investor that plays the square puzzle would see a square icon to replace in-group and the triangle icon to replace the word "out-group".

Conditional on each of these outcomes, and conditional on being matched with an in-group or out-group trustee, an investor would decide whether to trust or not trust in the second period. Thus, an investor would have up to eight strategy-method decisions to make in Period 2.[17]

Treatments

We studied two incentive conditions: *No punishment* and *Punishment*. These two conditions differed in the trust game stage. Under punishment, everything was identical to *No Punishment*, except that an investor who received $0 could decide whether to reduce $6 of the trustee's earnings in each period at no cost[18] to the investor (see Figure 2b). However, an investor could not know for certain whether this $0 was due to defection (the trustee chose Don't Roll) or if the trustee chose to Roll but was unlucky. An investor could not reduce the trustee's earnings if she chose OUT (and received $5) or if she chose IN and received $12 (i.e., the trustee chose to Roll and was lucky) (see sample Instructions in Appendix A).

Survey

After players finished their decisions, and before seeing the outcome, they were also incentivized to predict the decisions of their counterpart[19]. We asked investors: "How many of the trustees do you believe in Period 1 chose to ROLL for an A from their own (other) group?[20]" We asked trustees: "How many of investors do you believe in *Period 1* chose IN for a B from their own (other) group?" Participants were paid $1 for each correct answer. In the *Punishment* condition, we further asked trustees to predict how many of those A who had chosen IN would choose to

[17] Trustees in Period 2 also made decisions conditional on their possible Period 1 payoffs.

[18] We chose a costless rather than costly punishment, as we aimed to capture an environment where punishment is easy to implement and is salient to the one who might be punished. In many firms, for example, supervisors can easily "write-up" an employee for improper behavior, and doing so can have substantial negative consequences for the supervisee. Moreover, as our goal was to understand not only the impact of punishment but also how it might change depending on social identities, we wanted to ensure that punishment had the best chance to be used and was not deterred as a result of cost.

[19] Our belief elicitation is incentivized and occurs after subjects have finished their decisions. This procedure has been shown to produce accurate belief elicitations (Gachter and Renner 2010) and to avoid contamination of decisions during the experiment (Costa-Gomes and Weizsacker 2008). Despite this, it is possible that beliefs are impacted by decisions made in the game.

[20] In the experiment, trustees were referred to as being in Role B, while investors were in Role A. Players also received $1 if their answer was correct. We also reminded the players about the number of Role B's in the room.

punish if they receive zero. These answers helped us to determine whether investors' decision to punish changed with their beliefs regarding trustees' Roll (reciprocation) rates.

Experimental Procedure

We conducted all sessions at the Interdisciplinary Center for Economic Science laboratory using z-tree (Fischbacher 2007). Table 1 shows characteristics of the subjects who participated in the two conditions: these characteristics are not statistically significantly different (see Table B1 in Appendix B).

Upon arriving, subjects were seated in separate booths, so that they could neither see each other nor communicate before the experiment began. Before making decisions, subjects were randomly assigned an ID number, which determined their group and role in the group. Each group worked on the puzzle task in a different room with an experimenter present (standing at a distance). After the puzzle stage, all participants went back to their booths and all decisions in the trust game were made privately and anonymously through the computer interface. All sessions were finished within an hour. Subjects earned an average of $13, including a $5 show-up bonus.

Table 1: Summary statistics and comparison across treatments

	No Punishment	Punishment
Male	0.58	0.54
Caucasian	0.41[21]	0.40
African American	0.086	0.157
Asian	0.37	0.36
Hispanic	0.086	0.064
Other racial/Ethnic group	0.047	0.021
Observations	138	140

[21] All races reported under No Punishment are based on an observation of 128 rather than the entire data set 138 of the treatment due to missing reports.

4. Theoretical predictions

Trust decision

Assume investor *i* derives utility from earnings π_i as well as the trustee's earnings π_j[22]. If she chooses OUT (not to trust), both the investor and the trustee receive \$5 ($\pi_i = \pi_j = 5$), and the game ends. However, when she chooses IN, her expected payment is dependent upon the trustee *j*'s decision and, if the trustee chooses Roll, the outcome of that die roll.

Assuming the investor's utility is linear in her earnings and that of the trustee, we can express her expected utility of the trust decision as follows.

$$U_i = w_i\pi_i + w_j\pi_j$$

where $w_i \in [0,1]$ is the utility weight investor *i* assigns to her (expected) earnings π_i (henceforth, $E\pi_i$) and $w_j \in [0,1]$ is the weight *i* assigns to trustee *j*'s (expected) earnings, (henceforth $E\pi_j$)[23].

Let p_{ij} denote investor *i*'s subjective belief regarding trustee *j*'s likelihood of reciprocating (to Roll). Suppose that investor *i's* belief is a function of the absolute amount of punishment that *i* can impose upon the trustee *j* were he to "betray": $p_{ij} = f(|pun_{ij}|)$. Suppose further that p_{ij} monotonically increases in $|pun_{ij}|$, with $f' > 0, f'' < 0$. Note that in the *No punishment* condition $|pun_{ij}| = 0$, while in *Punishment* $|pun_{ij}| = 6$. Finally, we require that the belief follows $p_{ij} \in [0,1]$.

Investor *i* compares the expected utility she receives from choosing IN (see <1>) to the expected utility she receives from staying out (see <2>).

$$U_{IN}^i = w_i * E\pi_i + w_j * E\pi_j \qquad <1>$$

$$\text{where} \quad E\pi_i = p_{ij} * \left(\frac{1}{6} * 0 + \frac{5}{6} * 12\right) + \left(1 - p_{ij}\right) * 0 = 10p_{ij}$$

$$E\pi_j = p_{ij} * 10 + \left(1 - p_{ij}\right) * 14 = 14 - 4p_{ij}$$[24]

$$U_{OUT}^i = 5w_i + 5w_j \qquad <2>$$

[22] This utility set up is similar to that of Chen and Li (2009)'s group identity model.

[23] The weight parameter varies with identity and group boundaries. For example, an investor from the *Social* group may have an identical w_j for both in-group and out-group trustees, while an investor from the *Non-social* group could have a higher w_j for the in-group than out-group trustees.

[24] The specification of this model, which assumes that an investor cares both about self and the trustee would predict no punishment from the investor.

An investor chooses IN whenever $U_{IN}^{i} > U_{out}^{i}$.

Without loss of generality, set $w_i + w_j = 1$, meaning the condition for choosing IN is:

$$p_{ij} > \frac{14w_i - 9}{14w_i - 4}$$ [25]

We will refer to this belief $p_{thres} = \frac{14w_i - 9}{14w_i - 4}$, that leaves the investor indifferent between choosing IN and OUT, as the threshold belief.

Proposition 1: The threshold belief p_{thres} increases with w_i and decreases with w_j. That is, the relatively more (less) an investor cares about her own earnings and thus relatively less (more) about the other's earnings, the higher (lower) is the threshold belief required for an investor to choose IN.

We illustrate this effect in Figure 3a and 3b. Panel *a* assumes that an investor cares both about herself and about her counterpart: $w_i = 0.7$ and $w_j = 0.3$ so that $p_{thres} = \frac{14w_i - 9}{14w_i - 4} = 0.14$[26]. By contrast, panel *b* assumes that an investor only cares about herself, with $w_i = 1, w_j = 0, p_{thres} = \frac{14w_i - 9}{14w_i - 4} = 0.5$. Thus, the threshold belief increases from 0.14 to 0.5 after an increase in w_i.

Proposition 2: An investor will choose "Out" if her belief under no punishment is lower than the threshold belief and will choose IN if her belief under punishment is higher than the threshold belief.

This proposition implies that punishment is ineffective when beliefs under no punishment and punishment are on the *same* side of the threshold belief. By contrast, punishment promotes trust when these beliefs are on the *opposite* sides of the threshold belief. In particular, when the belief under no punishment is lower than the threshold belief, the model predicts the investor will

[25] This framework aims to explain the conditions under which an investor may trust or not trust based on her belief. It does not aim to explain why she holds that belief in the first place.

[26] Note that we chose these number to illustrate the basic idea of the model. It could well be the case that the utility function is not linear and the number would vary.

choose OUT, while the belief under punishment is higher than the threshold belief, the model predicts the investor to choose IN.

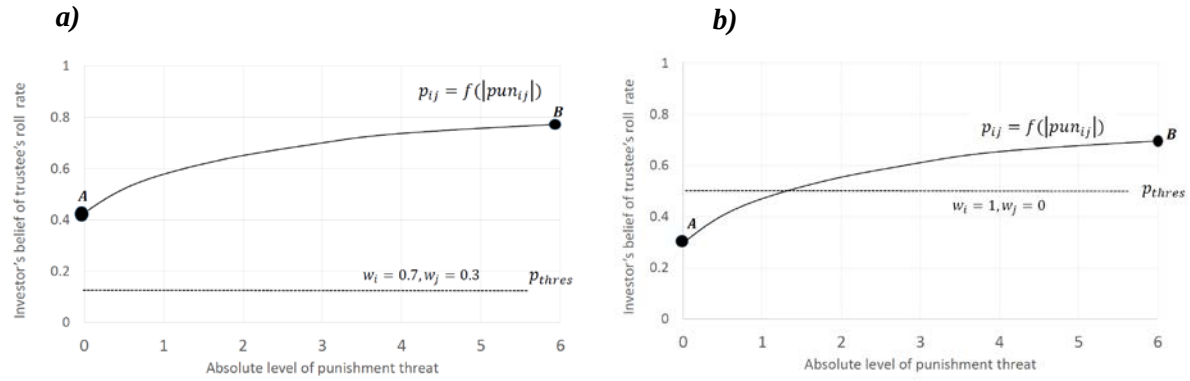


Figure 3. *a.* Punishment has *no effect* on changing the behavior of an investor who initially trusts. Points A and B represent the investor's belief regarding the trustee's roll rate under *No punishment* (A) and *Punishment* (B). The threshold belief is shown using a dashed line. This happens when the belief under no punishment (point A) and belief under punishment (point B) are on the *same* side of the threshold belief. ***b.* Punishment changes an investor's behavior from Out (no trust) to In (trust).** This happens when the belief under no punishment (point A) and belief under punishment (point B) are on the *opposite* sides of the threshold belief.

Predictions

In our experiment, *Social* investors experienced more exchange and sharing during the group formation process. If this results in broader group boundaries, as suggested by Buchan et al, then *Social* investors might be more willing to believe in both in- and out-group member's willingness to reciprocate absent punishment (e.g. Point A in panel 3*a*). Moreover, *Social* investors might also place greater weight on trustee earnings (an increase in w_j and a decrease in w_i), which implies a lower threshold belief to trust ($\frac{\partial p_{thres}}{\partial w_j} < 0$, see the dashed line in Figure 3*a* compared to that in 3*b*). As a result, both the beliefs with and without punishment are likely to be above the threshold level (on the *same* side of the threshold line), leaving punishment ineffective in changing trust decisions (Figure 3*a*, point A to B).

On the other hand, an investor from the *Non-social* group may draw relatively narrower group boundaries, and thus be less willing to believe in an out-group trustee's willingness to reciprocate. This belief may be even lower when punishment is absent (Figure 3*b*, point A). Further, they may put little weight on an out-group member's payoffs (corresponding to a

decrease in w_j and an increase in w_i), implying a higher threshold belief ($\frac{\partial p_{thres}}{\partial w_i} > 0$) that must be passed in order for a *Non-social* investor to trust an out-group member (see dashed line in Figure 3*b* in comparison to that in Figure 3*a*). Punishment increases the belief that an out-group trustee will reciprocate, perhaps increasing it above the threshold belief level. As a result, the beliefs under no punishment and that under punishment are likely to be on the *opposite* sides of the threshold belief level. This means that punishment may potentially change *Non-social* investors' decisions regarding whether to trust an out-group member (Figure 3*b*, point A to B).

5. Results

Our interests are in the effect of punishment on *Social* and *Non-social* investors' trust decisions and we are also interested in how *Social* and *Non-social* group investors punish., so we first focus on an investor's Period 1 trust and punishment decisions. At the end of this section, we report trustees' reciprocity decisions. In the discussion section, we offer thoughts on the mechanism underlying punishment decisions, and examine the effect of punishment opportunities on both trust and reciprocity.

5.1. Asymmetric effect of punishment on trust

Our results support the *Asymmetric Effect of Punishment on trust* hypothesis as described by the Propositions. We provide evidence first by comparing the trust frequencies between punishment and no punishment directly (see Figure 4) and we then use regressions to examine the effect of punishment after controlling for other variables (Table 2). We also show the robustness of this asymmetric effect of punishment by comparing the in-group favoritism between the two groups, both for the initial trust and after the investors are betrayed (Figure 5). We further explore the connection between belief and punishment by reporting beliefs when the environment includes or does not include punishment opportunities (Table 3). Finally, we provide regression evidence that the presence of punishment opportunities *per se* does not change behavior, but behavior changes can be explained entirely by the way punishment changes beliefs (Table 4).

RESULT 1: *After controlling for other factors, punishment increases Non-social group investors' trust only towards out-group trustees. Punishment has no effect on a Social investor's*

trust towards an out-group member. Punishment also has no impact on trust towards an in-group trustee for either group's investors.

Figure 4 reports that punishment promotes *Non-social* investors' trust towards out-group trustees (0.56 vs. 0.29, $p < 0.05$, two-sided Mann-Whitney test[27]), but not towards the in-group[28]. *Social* investors display no change in behavior as a result of punishment opportunities for either in-group or out-group trustees (In-group: 0.74 vs. 0.74, out-group: 0.59 vs. 0.59, $p > 0.1$). This result provides initial evidence for the "*Asymmetric effect of punishment on trust."*

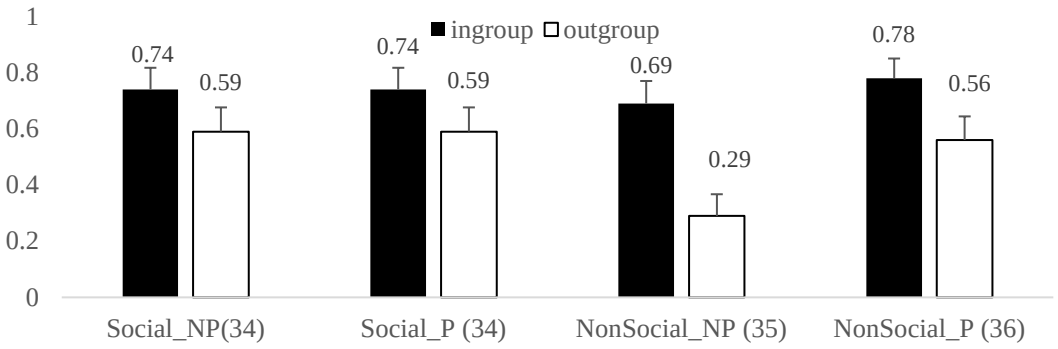


Figure 4 Trust frequency of investors in the *Social* and *Non-social* groups towards in-group and out-group trustees under *No Punishment* and *Punishment* conditions. P represents *Punishment*, NP stands for *No-punishment*. Numbers of observations are in parenthethes. The filled bar illustrates investors' trust towards in-group members. The open bar shows investors' trust towards out-group members.

Using an OLS regression controlling for the investor's group (a dummy for *Social* or *Not*), the punishment condition or not, and gender and session effects, we find punishment to significantly increase *Non-social* investors' trust towards out-group trustees (Column 5, Table 2, coefficient for punishment = 0.258, $p < 0.05$). We also find that being a *Social* investor promotes trust after taking into account the effect of punishment (Column 4, coefficient = 0.177, $p < 0.05$). We do not find punishment to change either investor's trust towards in-group trustees (Column 2 and 3, $p > 0.1$).

[27] Unless otherwise reported, all statistics are results from two-sided Mann-Whitney tests.

[28] To further identify the effect of punishment on trust, we ran two control treatments without the group identity formation process, one without punishment *Control_NP*, and one with punishment *Control_P*. Our results show punishment has a positive, but only marginal impact on promoting trust. The significant effect of punishment on a *Non-social group*'s willingness to trust is largely driven by substantial out-group discrimination when punishment is absent. These results are detailed in Appendix B.

Table 2: Effect of punishment on initial trust						
	(1)	(2)	(3)	(4)	(5)	(6)
		In-group			Out-group	
	All	Non-social	Social	All	Non-social	Social
Punishment	0.045	0.091	-0.025	0.127	0.258**	-0.020
	(0.076)	(0.098)	(0.116)	(0.079)	(0.118)	(0.112)
Social group	0.007			0.177**		
	(0.077)			(0.086)		
Male	-0.068	-0.142	0.021	0.072	0.131	0.055
	(0.084)	(0.136)	(0.126)	(0.088)	(0.140)	(0.126)
Session	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.912***	0.849***	0.997***	0.039	0.018	0.218
	(0.125)	(0.190)	(0.115)	(0.169)	(0.176)	(0.286)
Observations	139	71	68	139	71	68
R-squared	0.171	0.213	0.321	0.153	0.225	0.247

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

RESULT 2: *The asymmetric effect of punishment is robust to a diff-in-diff analysis, by comparing the in-group favoritism between the Social and the Non-social group. This effect also persists after an investor is betrayed.*

We conducted a diff-in-diff analysis between in- and out-groups, which shows in-group favoritism (trust towards an in-group "mirrors" trust towards an out-group) by *Social* and *Non-social* group members. In-group favoritism among the *Non-social* group is much larger than the *Social* group when punishment is absent (Figure 5*a*, mean = 0.4 vs. 0.15, $p < 0.05$, two-sided M-W test), while such favoritism difference disappears with punishment (Figure 5a, 0.22 vs 0.15, $p = 0.64$, two-sided M-W test). This pattern persists even after betrayal, where in-group favoritism under no punishment tends to be larger among the *Non-social* group than that under the *Social* group (Figure 5*b*, mean = 0.21 vs 0, $p < 0.1$, one-sided M-W test). In contrast, in-group favoritism between *Social* and *Non-social* group is not significantly different under punishment (mean = 0.11 vs 0, $p = 0.46$, two-sided M-W test). When betrayed by an out-group member, the bar continues to describe a similar trend, though the difference between the *Social* and *Non-social* group is not significant (Figure 5*c*, $p > 0.5$ for comparison under *No punishment* and *Punishment*).

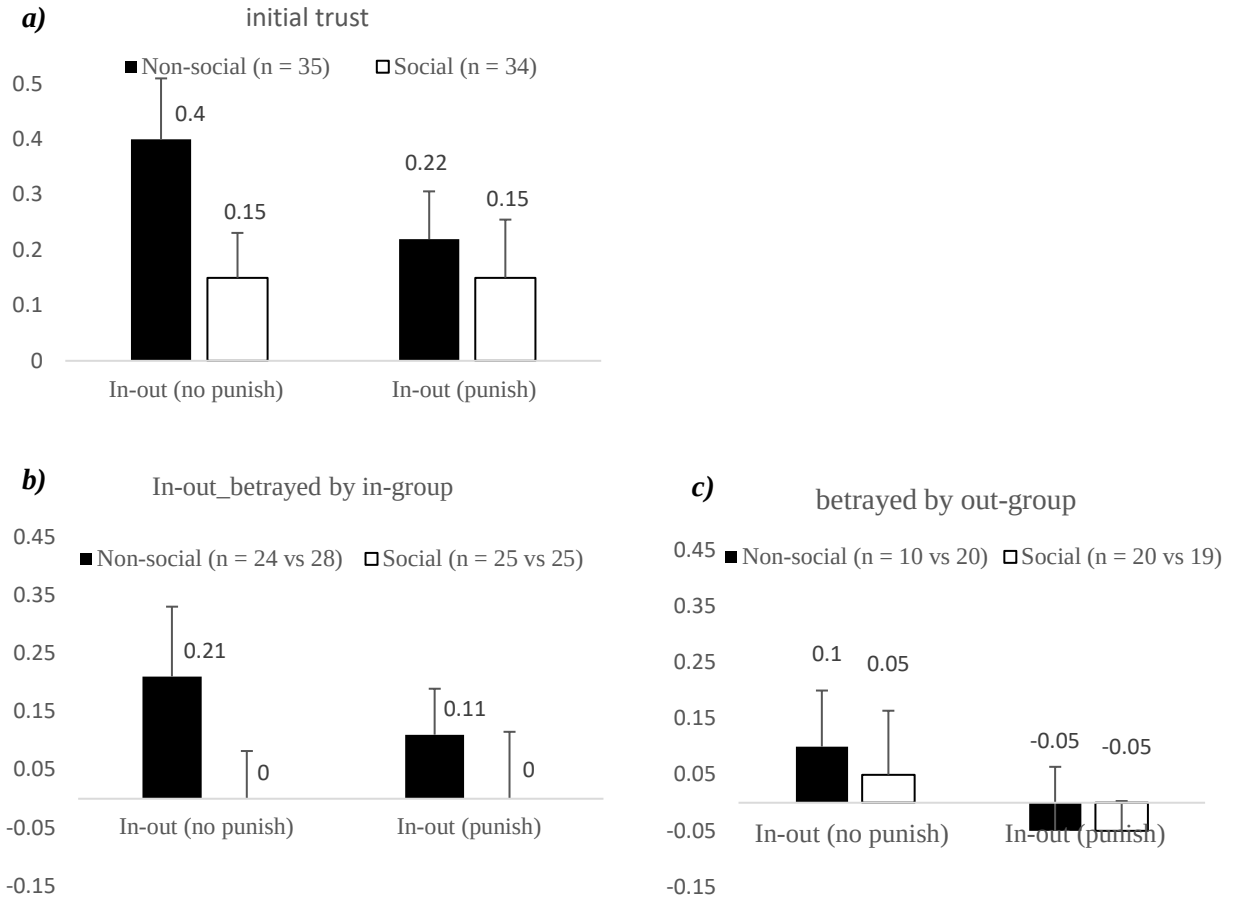


Figure 5: In-group favoritism by *Social* and *Non-social* groups. Each bar is calculated as the difference between one's trust towards in-group and that towards an out-group ("In-group – Out-group"), also, the in-group favoritism. The solid black bar represents the in-group favoritism of the *Non-social* group while the solid white bar represents that of the *Social* group. Panel *a* shows the in-group favoritism for the first time trust, while Panel *b* and Panel *c* show what happens if they are betrayed.

RESULT 3: *Social and Non-social investors' beliefs regarding in-group trustees' roll rates are independent of the presence of punishment. Both groups investors' beliefs about out-group trustees' reciprocation rates increase when punishment is present.*

We conducted an incentivized belief elicitation after investors completed their decisions, but before their results were revealed. In particular, after showing them the number of trustees in their session, we asked them: "How many of the trustees do you believe in *Period 1* chose to ROLL for an investor from their own (other) group?" With in-groups, punishment has no effect on the *Social* (0.65 vs 0.74) or the *Non-social* (0.60 vs. 0.65) investors' beliefs regarding

trustees' reciprocation rates ($p > 0.1$ for both, see Table 3)[29]. Yet, for both *Social* (0.47 vs 0.68) and *Non-social* (0.26 vs 0.49) investors, punishment contributes to significantly increasing investors' beliefs regarding out-group trustees' reciprocation rates ($p < 0.01$). We also compared the beliefs between the *Social* and *Non-social* investors. We find the beliefs of *Social* investors to be significantly higher than the *Non-social* investors only towards out-group trustees but both with (0.68 vs. 0.49, $p < 0.05$) and without punishment (0.47 vs. 0.26, $p < 0.01$). Yet, we do not find the belief of *Social* investors to be different from the *Non-social* investors towards in-group trustees both with (0.74 vs 0.65, $p > 0.1$) and without punishment (0.65 vs 0.60, $p > 0.1$).

Table 3: Investors' beliefs regarding trustees' reciprocation rates

	In-group			Out-group		
	Social (n=34, 34)	Non-social (n=35, 36)	M-W test	Social (n=34, 34)	Non-social (n=35, 36)	M-W test
No-Punishment	0.65	0.60	P = 0.41	0.47	0.26	P < 0.01
Punishment	0.74	0.65	P = 0.17	0.68	0.49	P < 0.05
M-W test	P=0.21	P=0.45		P < 0.01	P < 0.01	

RESULT 4: *After controlling for beliefs, punishment does not have any additional impact on trust decisions.*

Result 3 shows that investors believe reciprocity is more likely when punishment is possible. Result 4 reports a regression similar to that in Table 2, but with investors' beliefs of trustees' reciprocation rate added as a regressor. Beliefs are highly significant across all conditions ($p < 0.05$), but punishment is not significant for any condition (Table 4, $p > 0.1$) [30]. This suggests that the significant effect of punishment reported in Table 2 operates through the channel of beliefs.

[29] We assume that beliefs are monotonically increasing in the level of the punishment threat, with $f'>0$, $f''<0$. Results in Table 3 support this assumption.
[30] These results are robust to a probit regression analysis (See Table B2 in appendix B).

Table 4: Effect of punishment and belief on trust						
	(1)	(2)	(3)	(4)	(5)	(6)
		In-group			Out-group	
VARIABLES	All	Non-social	Social	All	Non-social	Social
Belief	0.590***	0.526**	0.652***	0.657***	0.594***	0.684***
	(0.115)	(0.238)	(0.177)	(0.133)	(0.203)	(0.141)
Punishment	-0.005	0.072	-0.103	-0.005	0.122	-0.139
	(0.073)	(0.099)	(0.107)	(0.084)	(0.128)	(0.114)
Social group	-0.039			0.055		
	(0.073)			(0.092)		
Male	-0.045	-0.071	0.027	0.091	0.136	0.103
	(0.077)	(0.133)	(0.123)	(0.085)	(0.125)	(0.131)
Session	Y	Y	Y	Y	Y	Y
Constant	0.543***	0.514	0.516***	-0.033	-0.138	0.123
	(0.204)	(0.326)	(0.170)	(0.173)	(0.212)	(0.333)
Observations	137	71	66	137	71	66
R-squared	0.306	0.294	0.481	0.295	0.319	0.395

Notes: The regression is based on the form: $Trust_{is} = \alpha + \beta Covariate_i + \delta_s + \epsilon_{is}$ where *Covariate* is listed to the left in the row and δ_s are the session fixed effects. The p-value is based on robust standard error in parentheses. *** p<0.01, ** p<0.05, * p<0.1

5.2 *Punishment behaviors*

RESULT 5*: Investors from the Non-social group use punishment significantly more frequently than their Social counterparts.*

Compared to their *Non-social* counterparts (n = 48, mean = 0.67), significantly fewer *Social* investors (n = 41, mean = 0.46) chose to reduce trustees' earnings when they received zero ($p < 0.1$, Figure 6). By controlling for whether one is from the *Social* or *Non-social* group, beliefs, investor's gender, whether the trustee is from the same group and session effects, we find an investor from the *Social* group is less likely to choose to use punishment[31] (see Table 4, Column 1, coefficient for being a social group member = - 0.314, $p < 0.01$).

This result remains significant after controlling for an investor's belief in a trustee's roll rate (Column 2, coefficient for membership = - 0. 235, $p < 0.01$). We further investigated the determinants of punishment by: 1) whether the trustee is from the same or other group (Columns

[31] This result is robust to using a probit regression ($p < 0.01$).

3 and 4). We find that belief plays no role in the decision to punish an in-group member. In contrast, an increasing belief in trustees' reciprocation rates contributes to reducing punishment towards out-group members (Column 4, coefficient = -0.623, $p < 0.05$). In Column 5- 10, we disaggregate the data by *Non-social* group (Column 5-7) or *Social* group (Columns 8-10). We find beliefs to play no role for a *Social* investor regardless whether it is an in-group or an out-group trustee ($p > 0.1$ for both, Column 9 and 10). We also find beliefs to have no effect on punishment towards in-group trustee among the *Non-social* investors ($p > 0.1$, Column 6). We find, however, that punishment decreases with increasing beliefs about the *out-group* trustees' reciprocation rates among the *Non-social* investors ($p < 0.05$, Column 7).

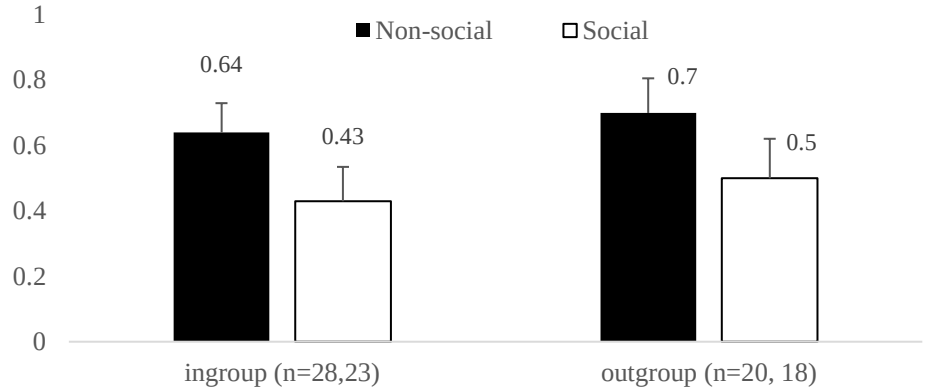


Figure 6: Frequency of punishment by Social and Non-social groups. Numbers in parentheses indicate the number of observations underlying each bar. The number above each bar indicates the mean frequency of punishment decisions.

Table 4: Determinants of Punishment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	All	All	In-group	Out-group	Non-social	NonSoc-In	NonSoc-Out	Social	Social-In	Social-Out
Belief		-0.478**	-0.402	-0.623**	-0.688**	-0.369	-0.863**	0.189	0.387	0.357
		(0.192)	(0.246)	(0.281)	(0.300)	(0.588)	(0.331)	(0.584)	(0.780)	(1.136)
Social group	-0.314***	-0.235**	-0.271*	-0.165						
	(0.116)	(0.112)	(0.137)	(0.134)						
Male	0.143	0.134	0.076	0.225	0.079	0.009	0.317	-0.058	0.038	-0.464
	(0.147)	(0.133)	(0.145)	(0.187)	(0.176)	(0.249)	(0.205)	(0.290)	(0.407)	(0.812)
In-group	-0.041	-0.003			0.061			-0.069		
	(0.073)	(0.074)			(0.120)			(0.136)		
Session	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	89	89	51	38	48	28	20	41	23	18
R-squared	0.338	0.386	0.460	0.534	0.491	0.501	0.804	0.427	0.627	0.563

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

5.3 Trustees' behavior

RESULT 6**:** *For both Social and Non-social trustees, the frequency of reciprocity towards in-group members is not impacted by punishment opportunities. Trustees are more likely to reciprocate to out-group investors when punishment opportunities exist.*

After controlling for whether there is punishment, whether one is from the *Social* or *Non-Social* group, the gender of the trustee, and using an OLS regression with robust standard errors clustered by subject, we find punishment to increase trustees' reciprocity towards out-group investors when there is an opportunity to punish ($p < 0.1$, Column 4, Table 5). Punishment does not impact the likelihood a trustee will reciprocate an in-group investor ($p > 0.1$, Column 1, Table 5). *Social* trustees do not reciprocate differently than the *Non-social* group trustees. If we control for trustees' beliefs regarding investors' likelihood to punish then we find the significance of punishment vanishes ($p > 0.1$, Column1, Table 6). Further, beliefs regarding the investor's likelihood to punish do influence *Non-social* trustees' reciprocity decisions ($p < 0.05$, Column 5, Table 6), but do not influence the *Social* group trustees ($p > 0.1$, Column 6, Table 6).

Table 5: Determinants of Reciprocal decisions

	(1)	(2)	(3)	(4)	(5)	(6)
		In-group			Out-group	
VARIABLES	All	Non-social	Social	All	Non-social	Social
Punishment	0.033	-0.116	0.176	0.164*	0.160	0.169
	(0.083)	(0.113)	(0.130)	(0.089)	(0.114)	(0.140)
Social group	-0.088			-0.005		
	(0.077)			(0.084)		
Male	-0.099	-0.095	-0.117	-0.017	-0.029	-0.076
	(0.092)	(0.123)	(0.126)	(0.094)	(0.142)	(0.121)
	(0.226)	(0.107)	(0.482)	(0.288)	(0.345)	(0.458)
Constant	0.956***	1.058***	0.800***	0.269	-0.080	0.578***
	(0.093)	(0.056)	(0.190)	(0.173)	(0.057)	(0.164)
Observations	135	66	69	135	66	69
R-squared	0.266	0.270	0.365	0.160	0.374	0.209

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Determinants of Reciprocal decisions after controlling for belief						
	(1)	(2)	(3)	(4)	(5)	(6)
		In-group			Out-group	
VARIABLES	All	Non-social	Social	All	Non-social	Social
believeAreduce6_percent	-0.086	0.084	-0.353	0.341	0.627**	0.189
	(0.177)	(0.251)	(0.332)	(0.227)	(0.244)	(0.435)
Punishment	0.067	-0.157	0.321	-0.018	-0.122	0.044
	(0.120)	(0.157)	(0.196)	(0.139)	(0.169)	(0.251)
Social group	-0.095			-0.022		
	(0.077)			(0.084)		
Male	-0.102	-0.092	-0.107	-0.054	-0.086	-0.087
	(0.094)	(0.127)	(0.127)	(0.097)	(0.132)	(0.135)
Constant	0.926***	1.047***	0.738**	0.161	0.022	0.391
	(0.126)	(0.069)	(0.306)	(0.160)	(0.079)	(0.317)
Session	Y	Y	Y	Y	Y	Y
Observations	133	66	67	133	66	67
R-squared	0.262	0.271	0.369	0.193	0.445	0.206

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6. Discussions and conclusions

6.1. Ineffectiveness of punishment on reciprocity

The insignificant difference between *Social* and *Non-social* trustees is consistent with the literature showing that trust is more easily manipulated than trustworthiness (Al-Ubaydli et al 2013). This may also be related to the finding that reciprocity is a social norm, while trust is not (Bicchieri et al. 2011). That we find punishment is often ineffective in increasing reciprocity rates is also in line with results from previous studies which find either ineffective or detrimental effects of sanctions on reciprocity (Fehr and Rockenbach 2003, Falk and Kosfeld 2006, Houser et al. 2008).

6.2. Explaining Punishment behavior

Punishment decisions in our environment are driven by social preferences: our experiment design ensures that punishment cannot be driven by strategic concerns[32]. Predictions of

[32] An investor can punish a potential violator in order to promote more positive future interactions with other in-group or out-group members. However, neither the trustee nor the investor learn their outcome until the end of the

prominent social preference models indicate one would choose *to punish* if she is inequity averse (Fehr and Schmidt 1999), independent of her belief regarding trustees' roll rates. In contrast, she would choose *not to punish* if she is efficiency driven and cares about the interests of the trustee, again independent of her beliefs (Andreoni and Miller 2002). The only scenario where one's punishment behavior is influenced by beliefs is when an investor has a preference for reciprocity (Dufwenberg and Kirchsteiger 2004). In this case she would be less likely to punish when she believes it is more likely that the trustee will reciprocate .

In Table 4, we find punishment behaviors to be independent of investors' beliefs when they are from the *Social* group (coefficient for "belief" > 0.1, Columns 8 – 10) or for both groups when punishment is towards an in-group member (Column 6 and 9, $p > 0.1$). This is consistent with behaviors of both inequity averse investors and efficiency driven investors. Yet, our results further showed that being a *Social* investor alone leaves one less likely to punish (Column 1, coefficient for "*Social* group", $p < 0.01$), suggesting that *Social* investors are more likely to be efficiency driven. This result is robust to controlling for beliefs over trustee's roll rates (Column 2, $p < 0.05$)

By contrast, we find that an increased belief in an out-group trustee's roll rate reduces a *Non-social* investor's likelihood to punish (Column 7, $p < 0.05$). This dependency of punishment on beliefs suggests that *Non-social* investors are reciprocal and punish trustees when they believe less in the likelihood of reciprocity. This is consistent with the prediction of the intention based reciprocity model (Dufwenberg and Kirchsteiger 2004).

6.3. Group boundaries

Our study extends the literature on group boundaries (Buchan et al. 2009, Rand et al. 2009). Our group formation process allows us to observe how *Social* and *Non-social* investors respond to the presence of punishment opportunities when it comes to interactions with in-group and out-group members. Moreover, while a bad outcome is certainly a result of a norm violation in traditional trust game (Berg et al 1995), here, a bad outcome can be purely due to bad luck

game and this is common knowledge. Therefore, the investor's punishment is more likely to be driven by social preference.

with good intentions. Thus, our results also shed light on how the group formation process helps to promote trust in a world with uncertainty and noise (Fudenberg Rand and Dreber 2010).

Our results also complement studies that investigate the underlying mechanism for unjustified blame towards a trustee who can have little control over a bad outcome (Baron and Hershey 1988, Gurdal et al, 2013, Rubin and Sheremeta 2015, Pan and Xiao 2016). Our paper provides another possible explanation for such blame; that is, depending on how their groups are formed, those with group norms that involve more sharing and exchange draw broader group boundaries, are more likely to treat an out-group counterpart similarly to an in-group counterpart, and are more forgiving towards those who may have defected.

6.4 Conclusion

Our study reveals that the way a group forms can impact its behavior, and in particular that "global" groups founded in social sharing and exchange activity can draw broader group boundaries than "local" groups that form in the absence of such pro-social activities. We developed a model, and empirical support for this model, showing that differences in beliefs underlie these different group boundaries. In particular, investors from *Social* groups are much more likely to believe that an out-group trustee will reciprocate than are investors from *Non-social* groups.

In our model, an investor has no incentive to punish a trustee even when a trustee's choice could lead the investor to receive zero. The reason is that utility is positively (at least non-negatively) impacted by both one's own and one's counterpart's earnings. Alternative social preferences, say involving envy, might imply that an investor's utility would be negatively impacted by trustees who receive positive earnings when an investor receives zero. While investigating this is outside the scope of our paper, it would be profitable for future research to explore both theoretically and experimentally the rich punishment and trust decisions that could emerge in such an environment.

A consequence of broader group boundaries is a reduced need or willingness to take advantage of punishment institutions to sanction those who may have defected. As punishment is costly, this can result in significant efficiency gains for global groups. Further, our results point

to the possibility of designing interventions to create opportunities for greater cooperation, sharing and exchange, among an existing group as part of broad-based policies to promote pro-social actions and attitudes.

Acknowledgements

We thank Tim Cason, Gary Charness, Yan Chen, David Eil, Charlie Plott for their helpful comments. We also thank seminar participants at ESA, University of California at Santa Barbara, Central South University, Nanyang Technological University, National University of Singapore, University of Pennsylvania.

References

Akerlof G. and Kranton., R (2005) Identity and the Economics of Organizations. *Journal of Economic Perspectives*, 19 (1): 9-32.

Al-Ubaydli O., Houser D., Nye J., Pagnelli, M., Pan, X. (2013) The causal effect of market participation on trust: An experimental investigation using randomized control. *PloSOne*. 8(3):e55968. doi:10.1371/journal.pone.0055968

Alesina A., Baqir, R. and Easterly, W. (1999) Public goods and ethnic divisions. *The Quarterly Journal of Economics*, 1243-84.

Andreoni J. and Miller J. (2002). Giving according to GARP: An experimental test of the consistency of preference for Altruism. *Econometrica*, 70 (2), 737-753

Baron, J and Hershey, J. (1988) Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, 54 (4), 569-579.

Berg, J., Dickhaut, J and McCabe. K. (1995) Trust, Reciprocity, and Social History. *Games and Economic Behavior*, 10, 122-142.

Bernhard, H., Fischbacher, U. and Fehr, E. (2006) Parochial altruism in humans. *Nature.* 444 (24), 912-915.

Bicchieri C., Xiao, E., Muldoon, R. (2011) Trustworthiness is a social norm, but trusting is not. *Politics, Philosophy & Economics.* 10 (2), 170-187.

Bottazzi, L, Darin, M. and Hellmann, T. (2016) The Importance of Trust for Investment: Evidence from Venture Capital. *Review of Financial Studies.* Doi:10.1093/rfs/hhw023.

Bohnet I., Herrmann B., Al-Ississ M., Robbett A., Al-yahya K. and Zeckhauser R. (2012) The elasticity of trust, in *Restoring trust in Organizations and Leaders: Enduring Challenges and Emerging Answers,* Roderick M. Kramer and Todd L. Pittinsky (eds.), New York: Oxfort University Press, 151- 169.

Buchan, N., Grimalda, G., Wilson, R., Brewer, M., Fatas, E. and Foddy, M. (2009) Globalization and human cooperation. *Proceedings of National Academy of Science (PNAS),* 106 (11): 4138-4142.

Charness, G., Rigotti L. and Rustichini A. (2007) "Individual Behavior and Group Membership". *American Economic Review* 97 (4). 1340-1352

Charness G. and Dufwenberg M. (2006) "Promises and Partnership." *Econometrica* 74, no. 6: 1579-1601.

Chen L., and Li X. (2009) Group Identity and Social Preferences. *American Economic Review* 99, no. 1: 431-457.

Costa-Gomes, M. and Weizsacker, G. (2008) *Stated Beliefs and Play in Normal-Form Games, Review of Economic Studies.* 75, 729 – 762.

Eckel C. and Grossman P. (2005) Managing diversity by creating team identity. *Journal of Economic Behavior & Organization* 58

Falk A. and Kosfeld, M (2006) The hidden costs of control. *American Economic Review,* 96 (5), 1611-1630.

Fehr, E. and Gachter, S. (2000) Cooperation and Punishment in Public Goods Experiment. *American Economic Review,* 90 (4), 980-994.

Fehr E., and Rockenbach, B. (2003) Detrimental effects of sanctions on human altruism. *Nature,* 422, 137-140.

Fehr, E. and Schmidt K. (1999) A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics,* 114 (3), 817-868.

Fudenberg D., Rand D. and Dreber A. (2012) Slow to anger and fast to forgive: Cooperation in an Uncertain World. *American Economic Review.* 102, 720-749.

Fukuyama F. (1995) *Trust: The Social Virtues and the Creation of Prosperity.* New York: Simon & Schuster.

Gachter, S., and Renner, E. (2010) *The effects of (incentivized) belief elicitation in public goods expeirments*. 13, 364 – 377.

Goette L, Huffman, D and Meier, S. (2006) The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups, *American Economic Review*, 96 (2): 212-216.

Meier, S., Goette, L, Huffman, D. and Sutter M. (2012) Competition between orgnizational groups: Its impact on altruistic and anti-social motivations. *Management Science,* 58 (5): 948-960.

Meier, S., Goette, L and Huffman, D (2012). "The Impact of Social Ties on Group Interactions: Evidence from Minimal Groups and Randomly Assigned Real Groups." *American Economic Journal: Microeconomics* 4, no. 1 : 101-115.

Gurdal, M., Miller, J. and Rustichini, A. (2013) Why Blame? *Journal of Political Economy,*121 (6), 1205-1247.

Herrmann et al. (2008) Antisocial punishment across societies. *Science*, 319, 1362

Herrmann B., Thoni C. and Gaechter, S. (2008) "Antisocial Punishment Across Societies." *Science* 319:1362-1367.

Houser D., Xiao E., McCabe, K., Smith, V. (2008) When punishment fails: Research on sanctions, intentions and non-cooperation. *Games and Economic Behavior*, 62 (2), 509-532.

Knack S, Keefer P (1997) Does social capital have an economic payoff? A cross-country investigation. Quarterly Journal of Economics 112: 1252–1288.

Lane T. (2016) Discrimination in the laboratory: A meta-analysis of economic experiments. *European Economic Review*, 90, 375-402.

Masella, P., Meier, S. and Zahn, P. (2014). Incentives and group identity. *Games and Economic Behavior*, 86, 12-25.

Mulder L., Dijk V., De Cremer D. and Wilke H. (2006) Undermining trust and cooperation: The paradox of sanctioning systems in social dilemmas. *Journal of Experimental Social Psychology,* 42, 147-162.

Okenfels, A. and Werner, P. (2014) Beliefs and ingroup favoritism. *Journal of Economic Behavior & Organization*, 108, 453-462.

Pan X. and Houser D. (2013) Cooperation during culture group formation promotes trust towards members of out-groups. *Proceedings of the Royal Society B,* doi: 10.1098/rspb.2013.0606

Pan X and Xiao E. (2016) It's not just the thought that counts: An experimental study on hidden cost of giving, *Journal of Public Economics*, 138, 22-31

Peysakhovich A., and Rand, D. (2016) Habits of Virtue: Creating Norms of Cooperation and Defection in the Laboratory, *Management Science,* 62 (3), 631-647.

Rand D., Pfeiffer T., Dreber A., Sheketoff R., Wernerfelt N. and Benkler Y. (2009) Dynamic remodeling of in-group bias during the 2008 presidential election. *Proceedings of National Academy of Science,* 106 (15), 6187-6191.

Rubin, J., Sheremeta., R (2016) Principal-agent Settings with Random Shocks. *Management Science*, 62 (4): 985-999.

Tajfel, H. and Turner, J. (1986) The Social Identity Theory of Intergroup Behavior. *In The Psychology of Intergroup Relations,* ed., Worchel, S. and Austin, W. 7-24. Chicago: Nelson-Hall.

Yamagishi T and Yamagishi M. (1994). Trust and Commitment in the United States and Japan. *Motivation and Emotion* 18 (2)

Yamagishi, T. (1986) The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology.* 51: 110-16.

Weng, Q. and Carlsson, F. (2015) Cooperation in teams: The role of identity, punishment and endowment distribution. *Journal of Public Economics.* 126: 25-38.

Zak P, Knack S (2001) Trust and growth. *Economic Journal*, 111: 295–321.

APPENDICES

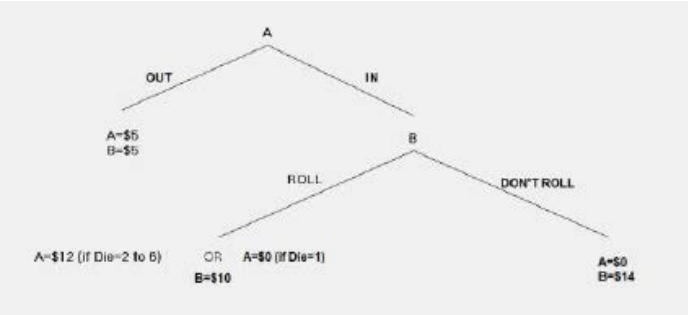
Appendix A

Hidden-effort trust game: Instructions (punishment)

<We presented these instructions on computer. Each Screen number below indicates the order of the screen on the computer. The subjects were able to navigate among screens as they wished.>

Screen 1.
In this game, you will keep the group membership you had in previous game.
There are two roles in your pair, one person will be randomly assigned the role of A, and the other will be assigned the role of B. The amount of money you earn depends on the decisions by you and your matched participant.

Screen 2.
Period 1. Person A chooses between IN and OUT. If A chooses OUT, then A and B each receives $5.
Next, each person B will choose between ROLL and DON't ROLL (a die). Note that B will not know whether A has chosen IN or OUT; however, since B's decision will only make a difference when A has chosen IN, we ask B's to presume (for the purpose of making this decision) that A has chosen IN.
If A has chosen IN and B chooses DON'T ROLL, then B receives $14 and A receives $0.
If B chooses ROLL, B receives $10 and the computer will roll a six-sided die to determine A's payoff. If the die comes up 1, A receives $0; if the die equals to 2, 3, 4, 5 or 6, A receives $12. This information is summarized in the chart below.



Both of you will be asked to make a decision:
For Person A: Will you choose IN or OUT?
For Person B: Will you choose ROLL or DON'T ROLL?

Screen 3
Before A knows B's decision, we give A the **chance to reduce** B's earnings conditional on **A's earnings.**

1. If A chose IN and received $0, A can **choose to reduce** B's earnings by **either 0 or $6.**
2. If A chose IN and received $12, then A **cannot reduce** B's earnings.

3. If A chose OUT, A receives $5 for sure and A **cannot reduce** B's earnings.

In a word, A **can reduce** B's earnings if A's earnings can be **influenced by B's decision** (indicate that A chose IN); and if **A receives $0** as a result of this decision. The reduction decision will be enforced if A received $0. The deducted amount **will not increase** A's earnings.

Screen 4

Period 2: In this period, you will play the **same** game as in Period 1, and with the **same** role (A or B) but with a new counterpart.

You **will not** know what happened in Period 1 when you make your decisions in Period 2. But in Period 2, you are able to make your decisions **baesd on possible outomce scenarios** you may have had in Period 1.

The next three pages will show you what it means to "**make decisions based on possible Period 1 outcome scenarios**."

Screen 5

For example, suppose you are assigned role A. In Period 1, you chose IN. When Period 2 starts, you will not know whether you have earned $0 or $12. Also, the role B participant will not know whether you chose IN or OUT, and therefore will not know their earnings as well.

In Period 2, you are still assigned role A. When you again choose between IN and OUT for the **new** counterpart, you will choose whether to decide IN or OUT if you received $12 in Period 1, and also wehther to decide IN or OUT if you received $0.

The next page will show you this example in detail.

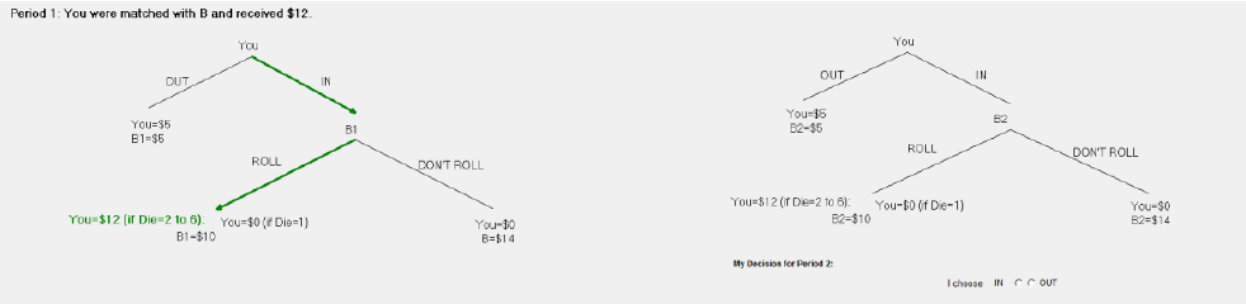
Screen 6

The left figure presents one of A's possible outcome in Period 1 when A chose IN: A received $12. The highlighted line indicates A chose **IN**. If **A received $12** as a result of this choice, then B must have chosen **ROLL,** thus is also highlighted (in solid line).

As A received $12 in this scenario, A is not provided the choice to reduce B's earnings.

In Period 2, **assuming $12 is A's Period 1 payoff**, we ask A the same question here for A's Period 2 counterpart (right figure)

Scenario 1



Screen 7

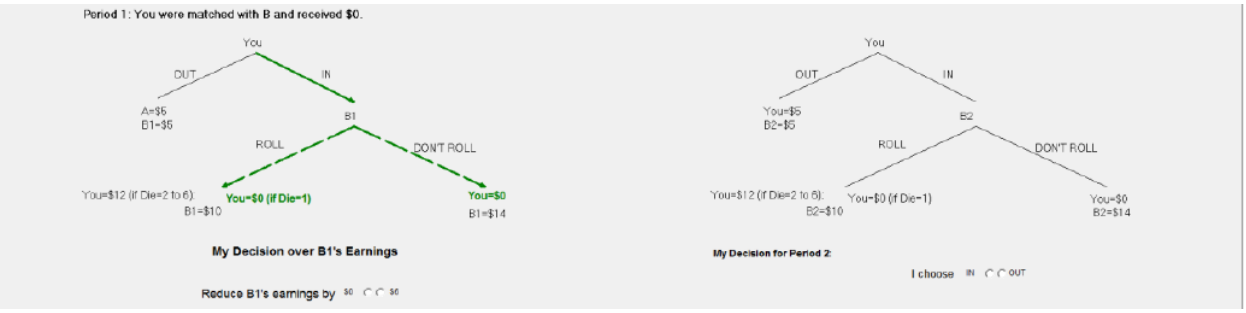
This page continue to present another possible outcome scenario when A chose IN: **If A received $0.**

Without knowing B's choice, either ROLL or DON'T ROLL is possible. So they are in dashed lines (left figure).

A can choose to reduce B's earnings by **either $0 or $6.**

A will be asked the same question in Period 2 A had in Period 1 **assuming A earned $0 in Period 1** (right figure)**.**

Scenario 2



Screen 8

Now let's go over Period 2 decisions for Person B. Suppose you are assigned role B. If you chose ROLL with die roll outcome bigger than 1. A will not be given the choice to reduce your earnings.

Or if you chose DON'T ROLL. Then if A chose OUT, both of you receives $5 regardless of your decision. A will not be given the choice to reduce your earnings.

Or if A chose IN, then A receives $0 and A will then be given the choice to reduce B's earnings.

Since B will not know A's reduction decision when B makes decision for his new counterpart in Period 2, we allow B to condition his decision on A's reduction decision.

1. **Assuming A reduced my earnings by $0 .**
2. **Assuming A redcued my earnings by $6.**

Next page will show you an example.

Screen 9

Suppose B chose DON'T ROLL. The left figure presents one of B's choices for A in Period 1. The highlighted line indicates B chose **DON'T ROLL.** If **B received $14** as a result of this choice, then A must have chosen **IN**, thus IN is also highlighted.

Meanwhile, A received $0 as a result of this decision. So A will be provided the choice to reduce B's earnings.

B will be asked the same question in Period 2 s/he had in Period 1 **assuming B earned $14** in Period 1.

B will make his decision conditional on A's reduction decision.

Screen 10

At the end of the experiment, either Period 1 or Period 2 will be randomly selected by the computer to determine your final payoff.

If Period 1 is selected, the decisions made by A and B in Period 1 determine their payoff.

If Period 2 is selectd, payoffs depend on the choices participants make in Period 2 under the corresponding outcome scenario that actually occurred in Period 1.

This is the END of the instruction. If you have any questions, please raise your hand and the experimenter will assit you.

Puzzle game instruction

Welcome. Today you will be participating in two experiments.

The first experiment is a puzzle game.

There is a sticker with your ID number on each of your screen. Please take if off and **attach** if on your shirt now. Half of you will have the TRIANGLE shaped sticker, and you belong to the TRIANGLE group; the other half of you will have the SQUARE shaped sticker belong to the SQUARE group.

There is also an envelope on each of your table. Please **do not** open the envelope now.

Now please stand up and look for your group members.

What's inside the envelope: Each of the envelopes contains FOUR pieces of cardboard.

Here is the task for each group:

1. The SQUARE group: to make SQUAREs.
2. The TRIANGLE group: to make TRIANGLEs.

The shape will be exactly the same as that of your sticker, but bigger.

The task will not be complete until each one of you:

1. The SQUARE group: make 5 squares of the same size.
2. The TRIANGLE group: make 5 triangles of the same size.

Group members are encouraged to share ideas and talk to each other during this exercise.

Note that when making shapes, the cardboard cannot overlap each other.

The winning team will receive $2 for each of their member.

If both of the teams were able to form their shapes, then the one who was faster will receive the prize. If both failed to make the shapes, then none of the teams will receive the prize.

You have 15 minutes to work on this task, when I start to time, you go to the work place now. If you finished, please raise your hand, the experimenter will check and will record the time your group have worked on the task.

Appendix B

Table B1: Demographic variables randomized in each condition

VARIABLES	(1) treatment	(2) treatment
Black	0.150	0.181
	(0.202)	(0.195)
Asian	-0.048	-0.091
	(0.156)	(0.147)
Hispanic	-0.281	-0.228
	(0.252)	(0.251)
Other	-0.274	-0.354
	(0.369)	(0.331)
Male	-0.034	-0.079
	(0.130)	(0.126)
attachment	-0.007	
	(0.026)	
Session dummies	Yes	Yes
	(0.129)	(0.119)
Constant	1.945***	1.933***
	(0.182)	(0.130)
Observations	258	268
R-squared	0.094	0.084

Notes: The regression is based on the form: Treatment = $Treatment_{is} = \alpha + \beta Covariate_i + \delta_s + \epsilon_{is}$ where *Covariate* is listed to the left in the row and δ_s are the session fixed effects

Caucasian has been dropped due to colinearlity. The p-value is based on robust standard error in parentheses. If we regress treatment status jointly on all covariates, we obtain a p-value for joint significance of 0.48 (without attachment) and 0.70 (with attachment)

Table B2: Determinants of belief on trust (Probit regression)

	(1)	(2)	(3)	(4)	(5)	(6)
		In-group			Out-group	
VARIABLES	All	Non-social	Social	All	Non-social	Social
believeb_roll_same_percent	2.103***	2.094**	2.734***			
	(0.482)	(0.860)	(0.780)			
believeb_roll_other_percent				2.094***	2.058***	2.569***
				(0.480)	(0.690)	(0.582)
Punishment	-0.026	0.326	-0.653	-0.034	0.334	-0.420
	(0.279)	(0.350)	(0.503)	(0.267)	(0.377)	(0.427)
Social group	-0.180			0.179		
	(0.298)			(0.282)		
Male	-0.172	-0.280	0.375	0.381	0.540	0.664
	(0.291)	(0.425)	(0.554)	(0.264)	(0.385)	(0.472)
Session effects	Y	Y	Y	Y	Y	Y
Constant	0.190	-0.028	-1.766**	-1.830***	-2.292**	-1.691*
	(0.740)	(0.932)	(0.736)	(0.619)	(0.948)	(0.907)
Observations	128	62	42	137	71	54

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The *Control* treatments

We ran two additional control treatments, one with punishment, *Control_P* and the other without *Control_NP* to help identify the effect of punishment on trust. That is, in the main result 1, we find punishment to promote a *non-social* group's trust towards out-group. But it is unclear, how much of this effect is due to the punishment effect alone (without group identity) and how much is due to its interaction effect with the group identity. A comparison between *Control_NP* and *Control_P* would identify the effect of punishment.

In the *Control* group (no group interaction before the trust game), punishment increases the willingness to trust, but this increase is not significant in comparison to the trust frequencies under the *No Punishment* (mean = 0.56) and *Punishment* conditions (mean = 0. 69, p = 0.31, two-sided M-W test). The significant effect of punishment on a *Non-social* group's willingness to trust is largely driven by substantial out-group discrimination when punishment is absent. The trust in *Control* (mean = 0.56) is significantly higher than the *Non-social* group's trust towards an out-group member (mean = 0.29, $p < 0.05$, two-sided M-W test). However, *Social* group formation leaves members less likely to discriminate against out-groups. In particular, the trust of a *Social-group* member in an out-group trustee is statistically identical to that of a *Control* group member (mean = 0.59 vs. mean = 0.56, p = 0.83, two-sided M-W test).

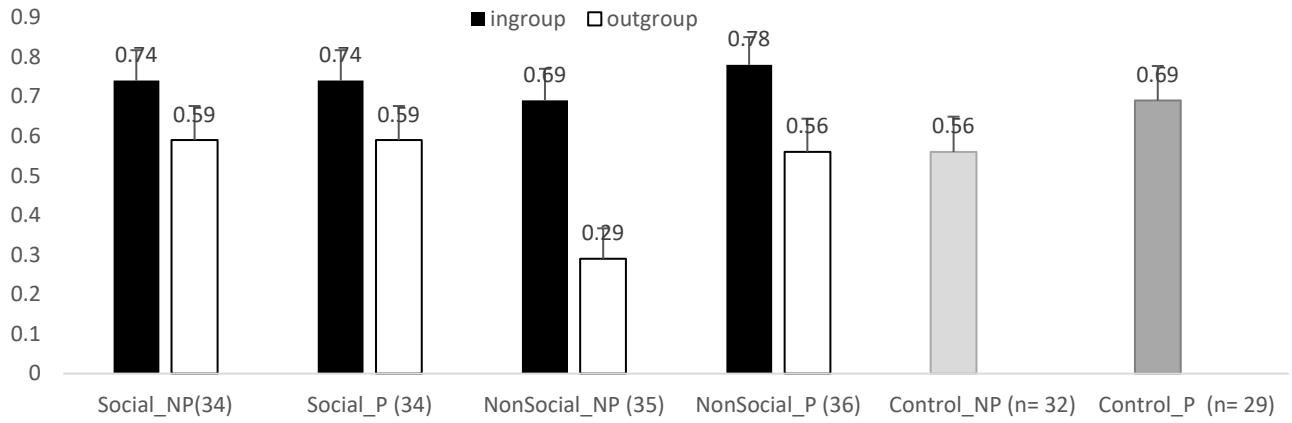


Figure A1: The Trust Frequency of *Social, Non-Social* and the *Control* groups. The grey bars show the trust of the control groups under No Punishment (*Control_NP*) and Punishment conditions (*Control_P*).