Choosing the agent's group identity in a trust game with delegated decision making*

Alexander Zimper† Nicky Nicholls‡

October 7, 2019

Abstract

Members of a given social group often favour members of their own group identity over people with different group identities. We construct a trust game in which the principal delegates the decision about an investment into a receiver to an agent who either favours the principal's or the receiver's group identity. When choosing the agent's group identity the principal faces a trade-off between a loyal agent and an agent who might increase the receiver's willingness to cooperate. We solve for the principal's decision in a subgame-perfect Nash equilibrium for the two scenarios of a risk-neutral and risk-averse agent, respectively.

Keywords: Trust Game; Identity; In-Group Bias; Fairness
JEL Classifications: C72, D82

*This paper benefitted from discussions with Sonali Das, Rangan Gupta, Hanjo Hamann, Jaromir Kovarik, Alexander Ludwig and Jesse Naidoo. We are particularly grateful for detailed comments and suggestions from an anonymous associate editor.

†Corresponding author. Department of Economics; University of Pretoria; Postal address: Private Bag X20; Hatfield 0028; South Africa; E-mail: alexander.zimper@up.ac.za

‡Department of Economics, University of Pretoria, South Africa. E-mail: nicky.nicholls@up.ac.za

1 Introduction

Starting with Tajfel (1970) there exists an extensive empirical literature at the intersection between psychology, sociology, and behavioral economics which demonstrates that considerations about group identity play an important role for individual preferences. A persistent finding in this literature are manifestations of an 'in-group bias' according to which an individual of a given social group tends to favour–or to trust–individuals of her own group identity over individuals that belong to a different social group (see, e.g., Tajfel and Turner (1986); Kuwabara et al. (2007); Chen and Li (2009); Sherif (2010); Balliet and van Lange (2013); Currarini and Mengel (2016); Morton et al. (2019)). Experiments that establish the existence of an 'in-group bias' in strategic situations typically take the form of dictator games–where one individual can allocate money to another individual–or of trust games–where one individual can invest money into another individual with the prospect of receiving some money in return.[1]

Applied to a situation in which a principal has to choose between employing either an agent of her own or of a different group identity, the existence of an 'in-group bias' seems to suggest that the principal will always favor the agent of her own over the agent of a different identity. However, this seemingly straightforward relationship between identity considerations and employment decisions does not capture the more complex–but relevant–strategic situations in which an agent with an in-group bias has to interact on behalf of the principal with a receiver who cares about the agent's group identity. Examples of such situations with delegated decision making that involve a receiver who is of a different group identity than the principal are aplenty. The original idea to our model goes back to a discussion about the historical question why the British East Indian Company (principal) might have chosen *gomasthas* (agents with local identity) rather than her own British employees to deal on her behalf with the local population (receiver). But there exist more mundane everyday situations where group identity considerations concerning the agent and the receiver are relevant to the principal and which, therefore, fall broadly under our modeling framework.[2]

[1]For early discussions of trust games see Berg et al. (1995) and references therein. For more recent studies see, e.g., Glaeser et al. (2000); Fershtman and Gneezy (2001); Ashraf et al. (2006); Burns (2006); Güth et al. (2008); Bornhorst et al. (2010); Chowdhury et al. (2016); Weisel and Zultan (2016); Daskalova (2018); Hamann and Nicholls (2018). Whereas Johnson and Mislin (2011) provide a meta-analysis for trust game experiments, Buchan et al. (2002) discuss the empirical evidence of an 'in-group bias' for a great variety of strategic set-ups, including dictator- as well as trust games.

[2]To quote an associate editor: "The white CEO who has staff that are mainly black – should he hire a white or black store manager. The charity trying to raise money in the US for disaster relief victims in Africa – who to hire as fundraisers. The microfinance bank that has rich managers but wants to appeal to lower income families – should she hire a rich or poor local representative. The white store

This paper develops a theoretical model about the possible impact of group identity considerations on the principal's employment decision in a strategic environment that involves the three players "principal", "agent", and "receiver", respectively.[3] The "principal" of a given group identity has to bankroll an investment into a "receiver" of a different group identity. The principal will only receive a positive return on this investment if the receiver cooperates towards the principal's objectives. We further assume that the principal has to delegate the decision about this investment to an "agent" whereby the principal cannot impose any enforceable contract about the agent's decision. As her only decision the principal can either choose an agent of her own or of the receiver's group identity. The choice of the agent's identity matters to the principal for two reasons. First, the agent has a strong in-group bias: whereas the agent with the principal's identity favors the principal, the agent with the receiver's identity favors the receiver. Second, by choosing the agent's identity, the principal might be able to influence the likelihood that the receiver will be of a cooperative type. In a nutshell, our model thus describes a situation in which the principal faces the trade-off between choosing an agent who is loyal to her versus an agent who is not loyal but might positively influence the receiver's willingness to cooperate.

Intuition suggests that the principal should choose the 'disloyal' agent with the receiver's identity only if this agent sufficiently increases the likelihood of a cooperative receiver. To make this intuition precise–and to go beyond it–we construct a trust game with delegated decision making for which we characterize the principal's choice as the outcome of a subgame-perfect Nash equilibrium (SPNE) in dependence on the model's payoff parameters. Recall that the original trust game is a two-stage game in which the principal decides about the amount a of money to be invested in the receiver. This amount becomes multiplied by a gross-return factor $k \geq 1$. In the second–and final–stage the receiver sends back some proportion λ of the gross return ka to the principal. We describe the principal as a risk-neutral expected utility maximizer who exclusively cares about material payoffs. In line with the experimental evidence from trust games, according to which receivers typically send back some money to the principal, we assume that the receiver is driven by fairness considerations.[4] To this purpose, we adopt the

manager trying to sell products to black or Asian customers – who to hire as store reps."

[3]To distinguish between these players, we refer throughout the paper to the "principal" as *she*, to the "receiver" as *he*, and to the "agent" as *it*.

[4]The behavioral economics literature on *fairness* is huge and can be, roughly, divided into two different strands. First–and relevant to our model–there is the *payoff distribution* approach according to which players care, in addition to their own material payoffs, about the allocation of other players' payoff by expressing, e.g., *inequality aversion* (cf., e.g., Fehr et al. 1993; Fehr et al. 1996; Fehr and Schmidt 1999; Bolton and Ockenfels 2000). Second, there is the *reciprocal/sequential kindness* approach according to which players care about the intentions with which they are treated by others (cf., e.g.,

proposed utility function in Fehr and Schmidt (1999) which measures the receiver's dislike of an advantageous inequality by some parameter $\theta \in [0,1]$: the greater the value of θ, the more the receiver dislikes the allocational unfairness that arises from him having a greater material payoff than the principal. We prefer to interpret θ as the cooperation type of the receiver whereby we distinguish between a non-cooperative type, denoted $\underline{\theta}$, and a cooperative type, denoted $\overline{\theta}$, such that $\underline{\theta} < \frac{1}{2} < \overline{\theta}$. In contrast to the cooperative receiver type, the non-cooperative type does not care much about any allocational unfairness arising for the principal. The receiver's cooperation type is private knowledge which is not observable by the principal.

As a modification of the original trust game, we assume that the principal has to delegate her investment decision to a third player, the agent, who decides about the amount a to be sent to the receiver. For reasons that lie outside of the model, the agent cannot be committed by the principal to any course of action. The principal's only decision is therefore the choice between the agent's two possible group identities: the principal's identity, denoted $\mathbf{P}$, or the receiver's identity, denoted $\mathbf{R}$, respectively. For a given common prior μ defined over the receiver's type space, we denote by $\mu\left(\overline{\theta} \mid \mathbf{P}\right)$ and $\mu\left(\overline{\theta} \mid \mathbf{R}\right)$ the conditional probabilities that the receiver is of the cooperative type when the principal chooses an agent with identity $\mathbf{P}$ or $\mathbf{R}$, respectively. Through these conditional probabilities we formalize the notion that the agent's group identity has an impact on the receiver's willingness to cooperate with the principal. Central to our analysis will be the inequality

$$\mu\left(\overline{\theta} \mid \mathbf{P}\right) < \mu\left(\overline{\theta} \mid \mathbf{R}\right) \tag{1}$$

according to which the probability of a cooperative receiver type is greater for an agent who shares the receiver's group identity than for an agent who shares the principal's group identity. Simply put, (1) means that the receiver's willingness to cooperate is greater when the agent is of his own group identity.

The agent of our model expresses a strong in-group bias by favoring members of its own identity group. For the agent with the receiver's identity we simply model this in-group bias through a utility function which coincides with the receiver's utility function. For the agent with the principal's identity we distinguish between two different scenarios. In a first scenario, we consider an agent who has the same risk-neutral expected utility function as the principal. In a second scenario, we consider a risk-averse agent who maximizes the expected utility over the principal's material payoffs such that the agent's Bernoulli utility function becomes infinitely steep when the payoffs approach zero. We comprehensively characterize the principal's choice of the agent's identity as the outcome

Rabin 1993; Dufwenberg and Kirchsteiger 2004; Falk and Fischbacher 2006).

of the subgame-perfect Nash equilibrium of the game (which is essentially unique). For the first scenario of a risk-neutral agent we identify the range of payoff parameter values such that the inequality (1) is necessary and sufficient for the principal to choose the agent with the receiver's identity (cf. Theorem 1).[5] This first scenario thus confirms our original intuition according to which the principal would only choose the 'disloyal' agent with the receiver's identity if this choice leads to a greater willingness to cooperate from the receiver. More precisely, for a relevant range of payoff parameters the principal chooses the agent with the receiver's identity over the loyal agent of her own identity if and only if this choice comes with a greater likelihood of a cooperative receiver.

The situation is different for our second scenario which drives a wedge between the principal's and the loyal agent's risk preferences (cf. Theorem 2). Now there exist situations in which the principal strictly prefers the agent with the receiver's identity $\mathbf{R}$ over the agent with her own identity $\mathbf{P}$ even if the inequality (1) is violated. The key to this finding lies in the fact that the risk-neutral principal of our model will always prefer either a maximal or a zero-investment. A loyal agent who strongly abhors the possibility of a zero payoff, however, will never go for the maximal investment amount. In such situations, the principal might, by a continuity argument, strictly prefer the agent with the receiver's identity over the loyal agent of her own identity even if this choice (slightly) decreases the likelihood of a cooperative receiver. We make this argument precise for the special case of a risk-averse agent with logarithmic Bernoulli utility function.

The remainder of the paper proceeds as follows. Section 2 constructs the game, which is solved for the first scenario of a risk-neutral agent in Section 3. Section 4 analyzes the second scenario of a risk-averse agent. Section 5 concludes. Formal proofs are relegated to the Appendix.

2 Trust game with delegated decision making

2.1 Game structure

There are three players: the principal, the agent, and the receiver. The principal and the receiver have, by assumption, different group identities. The receiver can be of two different cooperation types: he is either cooperative ($\overline{\theta}$) or non-cooperative ($\underline{\theta}$) from the perspective of the principal. The trust game with delegated decision making consists of

[5] To be precise, the model's payoff parameters have to satisfy the threshold condition $k\mu\left(\overline{\theta} \mid \mathbf{R}\right) > 2$ according to which the expected return from choosing $\mathbf{R}$ is sufficiently large.

three different stages at which strategic decisions are made as well as a chance move by nature following the principal's choice of the agent's identity.

2.1.1 First stage

The principal chooses either an agent of her own ($\mathbf{P}$) or of the receiver's ($\mathbf{R}$) group identity. While doing so, the principal cannot observe the receiver's cooperation type. A strategy of the principal, denoted s_p (or simply i for *identity*), is thus simply either $\mathbf{P}$ or $\mathbf{R}$.

2.1.2 Chance move by nature

Conditional on the principal's choice of the agent's identity nature randomly determines the receiver's cooperation type $\theta \in \left\{\underline{\theta}, \overline{\theta}\right\}$. Denote by $\mu(\theta \mid i)$ the conditional probability that the receiver will be of type θ when the agent's identity is $i \in \{\mathbf{P}, \mathbf{R}\}$.

It seems natural that the receiver is strictly more likely to cooperate whenever the agent is of his own identity $\mathbf{R}$, i.e., $\mu\left(\overline{\theta} \mid \mathbf{P}\right) < \mu\left(\overline{\theta} \mid \mathbf{R}\right)$. However, there are also perceivable situations where it is plausible to observe $\mu\left(\overline{\theta} \mid \mathbf{P}\right) = \mu\left(\overline{\theta} \mid \mathbf{R}\right)$ or even $\mu\left(\overline{\theta} \mid \mathbf{P}\right) > \mu\left(\overline{\theta} \mid \mathbf{R}\right)$. The latter case may arise if the receiver particularly despises agents of his own identity as 'traitors' or 'enemy collaborators'. The former case corresponds to situations in which the receiver simply does not care about the identity of the 'middleman' but only about the amount invested into him. Our analysis imposes as only restriction that these conditional probabilities are non-degenerate, i.e., we only assume that

$$0 < \mu\left(\overline{\theta} \mid \mathbf{P}\right), \mu\left(\overline{\theta} \mid \mathbf{R}\right) < 1$$

2.1.3 Second stage

The agent observes its identity $i \in \{\mathbf{P}, \mathbf{R}\}$ chosen by the principal. If the agent has the same identity as the principal, $\mathbf{P}$, it cannot observe the receiver's true cooperation type. In contrast, the agent with the receiver's identity, $\mathbf{R}$, knows the true cooperation type of the receiver. Although this difference in the information of both agent types is not substantial to our results, we feel that it is more appropriate whenever people of the same identity have 'a better understanding of each other'.[6] The agent has to allocate an amount $a \in [0, 1]$ to the receiver. Our preferred interpretation is that a stands for a social investment in the receiver paid for by the principal. A strategy of the

[6] In a previous version of this paper we assumed that both agent types could not observe the cooperation type of the receiver. While the equilibrium outcomes were basically the same, the notation became significantly more complicated.

agent, denoted s_A, assigns to every possible observation $\mathbf{P}$ and $\{\mathbf{R},\theta\}$, $\theta \in \{\underline{\theta},\overline{\theta}\}$, an amount $a \in [0,1]$. We write $s_A(\mathbf{P})$ for the action that strategy s_A assigns to the agent of identity $\mathbf{P}$ and $s_A(\mathbf{R},\theta)$ for the action that s_A assigns to the agent of identity $\mathbf{R}$ who has observed the cooperation type $\theta \in \{\underline{\theta},\overline{\theta}\}$.

2.1.4 Third stage

The receiver observes the agent's identity $i \in \{\mathbf{P},\mathbf{R}\}$, the amount $a \in [0,1]$ allocated to him, as well as his own cooperation type $\theta \in \{\underline{\theta},\overline{\theta}\}$. The receiver gets awarded a multiplied payment ka such that the gross-return parameter k satisfies $k \geq 1$. The receiver has to decide about the proportion $\lambda \in [0,1]$ of the total amount ka that will be sent to the principal. Our preferred interpretation is that the amount λka stands for the principal's return on the social investment a. A strategy of the receiver, denoted s_R, assigns to every possible observation

$$(i,a,\theta) \in \{\mathbf{P},\mathbf{R}\} \times [0,1] \times \{\underline{\theta},\overline{\theta}\}$$

some proportion $\lambda \in [0,1]$. We write $s_R(i,a,\theta)$ for the action that strategy s_R assigns to the receiver of cooperation type $\theta \in \{\underline{\theta},\overline{\theta}\}$ who has observed the agent's identity $i \in \{\mathbf{P},\mathbf{R}\}$ and the agent's action $a \in [0,1]$.

2.2 Payoffs and utilities

2.2.1 Principal

For fixed actions a and λ of the agent and the receiver, respectively, the principal's material payoff is given as

$$\pi_P(a,\lambda) = 1 - a + \lambda ka \tag{2}$$

The principal is a risk-neutral expected utility maximizer who only cares about her expected material payoffs. Fix the strategies (s_A, s_R) of the agent and the receiver, respectively. The principal's expected utility from choosing the agent with her own identity $\mathbf{P}$ is then given as

$$U_p(\mathbf{P},s_A,s_R) = \sum_{\theta\in\{\underline{\theta},\overline{\theta}\}} \left(1 - s_A(\mathbf{P}) + s_R(\mathbf{P},\theta,s_A(\mathbf{P}))\, ks_A(\mathbf{P})\right)\mu(\theta \mid \mathbf{P}) \tag{3}$$

If the principal chooses instead the agent with the receiver's identity $\mathbf{R}$, her expected utility becomes

$$U_p(\mathbf{R},s_A,s_R) = \sum_{\theta\in\{\underline{\theta},\overline{\theta}\}} \left(1 - s_A(\mathbf{R},\theta) + s_R(\mathbf{R},\theta,s_A(\mathbf{R},\theta))\, ks_A(\mathbf{R},\theta)\right)\mu(\theta \mid \mathbf{R})$$

2.2.2 Receiver

In line with the literature on fairness we assume that the receiver cares about a combination of (i) the principal's material payoff (2) and (ii) his own material payoff which is, for fixed actions a and λ, given as

$$\pi_R(a,\lambda) = (1-\lambda)\,ka$$

More specifically, we now assume that the cooperation type θ of the receiver takes on some numerical value in the unit interval whereby we fix the values of $\underline{\theta}$ and $\overline{\theta}$ such that $\underline{\theta} \in \left[0, \frac{1}{2}\right)$ and $\overline{\theta} \in \left(\frac{1}{2}, 1\right]$. This assumption allows us to define the utility of a receiver of cooperation type $\theta \in \left\{\underline{\theta}, \overline{\theta}\right\}$ for fixed actions a and λ by a piecewise-linear utility function adopted from Fehr and Schmidt (1999, equation 1):

$$U_R(a,\lambda;\theta) = \begin{cases} \pi_R(a,\lambda) - \theta\left(\pi_R(a,\lambda) - \pi_P(a,\lambda)\right) & \text{if } \pi_R(a,\lambda) \geq \pi_P(a,\lambda) \\ \pi_R(a,\lambda) - \left(\pi_P(a,\lambda) - \pi_R(a,\lambda)\right) & \text{else} \end{cases} \tag{4}$$

If the difference $\pi_R(a,\lambda) - \pi_P(a,\lambda)$ between receiver's and principal's material payoffs is strictly negative, the fairness literature speaks of a "disadvantageous inequality". The greater this "disadvantageous inequality" the smaller becomes the receiver's utility. Consequently, if the receiver is allocated a positive amount $a > 0$ we will never observe a choice λ^* such that the receiver gets a strictly smaller payoff than the principal, i.e., we cannot have that $\pi_R(a,\lambda^*) < \pi_P(a,\lambda^*)$ whenever $a > 0$.

Focus therefore on the case $\pi_R(a,\lambda) \geq \pi_P(a,\lambda)$ where the receiver's payoff is (weakly) greater than the principal's payoff. A strictly positive difference $\pi_R(a,\lambda) - \pi_P(a,\lambda)$ stands for an "advantageous inequality". If the receiver has fairness considerations described by (4), such an "advantageous inequality" comes with a utility-loss that is weighted by the receiver's cooperation type: for the non-cooperative type $\underline{\theta}$ the loss from an "advantageous inequality" matters less than for the cooperative type $\overline{\theta}$.[7]

2.2.3 Agent

This paper considers agents who are extremely biased towards their own group identity. To model the in-group bias of the agent with the receiver's identity **R**, we simply assume that this agent has the same utility function as the receiver. That is, for a fixed strategy s_R of the receiver satisfying $s_R(\mathbf{R}, a, \theta) = \lambda$ the agent's utility from action a is

$$U_A(a, s_R; \mathbf{R}, \theta) = U_R(a,\lambda;\theta) \tag{5}$$

[7] The fairness literature typically assumes that an "advantageous inequality" results in a smaller utility loss than a "disadvantageous inequality" of the same size. For the utility function (4) this assumption is equivalent to $\theta \leq 1$ which is satisfied for the cooperation parameter values of our model.

such that $U_R(a, \lambda; \theta)$ is given by (4).

Turn now to the agent with the principal's identity $\mathbf{P}$ who cannot observe the receiver's cooperation type. We assume, in a first scenario, that the preferences of principal and agent are perfectly aligned.[8] That is, the agent's expected utility coincides with the principal's expected payoff (3): for a fixed strategy s_R of the receiver, the expected utility of the agent with identity $\mathbf{P}$ from action $a \in [0, 1]$ is given as

$$U_A(a, s_R; \mathbf{P}) = \sum_{\theta \in \{\underline{\theta}, \overline{\theta}\}} (1 - a + s_R(\mathbf{P}, a, \theta) ka) \mu(\theta \mid \mathbf{P})$$

3 Solving the game

We solve the game for a subgame-perfect Nash equilibrium whereby we proceed by backward induction.

3.1 Receiver

In the final stage, the receiver of cooperation type $\theta \in \{\underline{\theta}, \overline{\theta}\}$ decides about the amount of money he sends back to the principal after he has observed the identity $i \in \{\mathbf{P}, \mathbf{R}\}$ and the choice $a \in [0, 1]$ of the agent. Technically speaking, every observation

$$(i, a, \theta) \in \{\mathbf{P}, \mathbf{R}\} \times [0, 1] \times \{\underline{\theta}, \overline{\theta}\}$$

stands for a *node* at which a subgame starts which simply corresponds to the receiver's decision.

If the receiver has been sent the zero amount $a = 0$, both receiver types are completely indifferent between all their possible actions on the unit interval. To keep our analysis simple, we discard non-interesting cases of multiple best responses and focus on the best response $f_R(0, \theta) = 0$ for both types $\theta \in \{\underline{\theta}, \overline{\theta}\}$. Consider now on the case $a > 0$. Even the cooperative type of the receiver who cares about fairness will not send back a greater amount to the principal than he keeps for himself. This gives us an upper boundary for the receiver's possible best responses with zero being the lower boundary. The next proposition (formally proved in the Appendix) uses the fact that the receiver's utility function $U_R(a, \lambda; \theta)$ is strictly increasing for $\theta > \frac{1}{2}$ whereas it is strictly increasing for $\theta < \frac{1}{2}$ on the relevant interval of possible best responses. To keep the notation simple, we write

$$s_R^*(a, \theta) \equiv s_R^*(\mathbf{R}, a, \theta) = s_R^*(\mathbf{P}, a, \theta)$$

[8]A second scenario will be discussed in Section 4 where we consider a 'misalignment' between the risk-preferences of the principal and the agent with the principal's identity.

for the receiver's SPNE strategy because the receiver's utility does not depend on the principal's choice of the agent's identity.

Proposition 1.

(a) *The SPNE action of the non-cooperative receiver type* $\underline{\theta}$ *is given as* $s_R^*(a, \underline{\theta}) = 0$.

(b) *The SPNE action of the cooperative receiver type* $\overline{\theta}$ *is given as*

$$s_R^*\left(a, \overline{\theta}\right) = \begin{cases} 0 & \text{if } a \leq \frac{1}{1+k} \\ \frac{(1+k)a-1}{2ka} & \text{if } a \geq \frac{1}{1+k} \end{cases} \tag{6}$$

Observe that the share $s_R^*\left(a, \overline{\theta}\right)$ of the total cake that the cooperative receiver type will send back to principal depends on the parameter k which determines the size of the cake. To see why, suppose that the cooperative receiver type has been sent the amount $a \in \left[\frac{1}{1+k}, 1\right]$. Because this receiver type cares about fairness, he wants to minimize the difference between his and the principal's payoffs so that the optimal share λ^* is pinned down by the equation

$$\begin{aligned} \pi_R(a, \lambda^*) &= \pi_P(a, \lambda^*) \\ &\Leftrightarrow \\ \frac{(1+k)a-1}{2ka} &= \lambda^* \end{aligned}$$

This share is strictly increasing in a on the interval $\left[\frac{1}{1+k}, 1\right]$ reaching with $\frac{1}{2}$ its maximum at $a = 1$. For smaller investment amounts $a \in \left[\frac{1}{1+k}, 1\right)$, however, the cooperative receiver sends back a strictly smaller share than $\frac{1}{2}$, which can be seen from the following inequality

$$\begin{aligned} \frac{(1+k)a-1}{2ka} &< \frac{1}{2} \\ &\Leftrightarrow \\ a &< 1 \end{aligned}$$

3.2 Agent

Following the chance move of nature the agent with identity **R** observes the receiver's cooperation type $\theta \in \left\{\underline{\theta}, \overline{\theta}\right\}$. This agent has to choose SPNE actions for the two subgames starting at the nodes $(\mathbf{R}, \underline{\theta})$ and $\left(\mathbf{R}, \overline{\theta}\right)$, respectively. In contrast, the agent with the same identity **P** as the principal does not observe the receiver's cooperation type. This agent has to choose a SPNE action for the subgame starting with the (unobserved) chance move by nature after the principal has chosen identity **P**.

3.2.1 Agent with identity R

By assumption, the agent who has the same identity $\mathbf{R}$ as the receiver has also the same utility as the receiver. Substituting the receiver's SPNE strategy from Proposition 1 into the agent's utility (5) results in

$$\begin{aligned} U_A\left(a, s_R^*\left(a,\theta\right); \mathbf{R}, \theta\right) &= \pi_R\left(a, s_R^*\left(a,\theta\right)\right) - \theta\left(\pi_R\left(a, s_R^*\left(a,\theta\right)\right) - \pi_P\left(a, s_R^*\left(a,\theta\right)\right)\right) \\ &= \left(1-\theta\right)\left(1 - s_R^*\left(a,\theta\right)\right) ka + \theta\left(1 - a + s_R^*\left(a,\theta\right) ka\right) \qquad (7) \end{aligned}$$

provided that the receiver's payoff is (weakly) greater than the principal's payoff. Intuitively, we would expect that an agent who shares the receiver's utility function is going to send the maximal amount $a = 1$ to the receiver. The following proposition (formally proved in the Appendix) confirms this intuition–up to a case of indifference for $k = 1$–for the utility specification (7).

Proposition 2. *Fix the receiver's SPNE strategy s_R^*. The SPNE action of the agent with identity $\mathbf{R}$–which is the same for both cooperation types of the receiver–is given as*

$$s_A^*\left(\mathbf{R}\right) \equiv s_A^*\left(\mathbf{R}, \underline{\theta}\right) = s_A^*\left(\mathbf{R}, \overline{\theta}\right) = 1$$

Proposition 2 seems to be straightforward. And indeed, sending the maximal amount $a = 1$ is the unique best response of the agent with identity $\mathbf{R}$ whenever (i) the receiver type is non-cooperative or (ii) the receiver type is cooperative and the return parameter satisfies $k > 1$. The situation is more complex, however, when the receiver type is cooperative but $k = 1$. In this case, the overall size of the 'cake' is not increased by sending any money to the receiver. If the cooperative type received, for example, the maximal amount $a = 1$ from the agent, he would therefore maximize his utility (4) by sending half of this amount back to the principal so that principal and receiver end up with material payoff $\frac{1}{2}$. This optimal allocation from the cooperative receiver's perspective could be alternatively obtained if the agent only sent amount $a = \frac{1}{2}$ to the receiver who then keeps the whole amount. Similarly, any amount $a \in \left(\frac{1}{2}, 1\right)$ sent to the receiver would result in the allocation where principal and receiver end up with material payoff $\frac{1}{2}$.[9]

[9] Again, in order to avoid keeping track of non-relevant indifferences, we restrict attention to the unique SPNE action specified in Proposition 2.

3.2.2 Agent with identity P

Given the SPNE strategy s_R^* of the receiver, the expected utility of the risk-neutral agent with identity $\mathbf{P}$ from choosing a is given as

$$U_A(a, s_R^*; \mathbf{P}) = \sum_{\theta \in \{\underline{\theta}, \overline{\theta}\}} (1 - a + s_R^*(a, \theta)\, ka)\, \mu(\theta \mid \mathbf{P}) \tag{8}$$

which becomes after substituting (6) for $s_R^*(\overline{\theta}, a)$:

$$U_A(a, s_R^*; \mathbf{P}) = \begin{cases} 1 - a & \text{if } a \leq \frac{1}{1+k} \\ 1 - a + \frac{(1+k)a - 1}{2} \mu(\overline{\theta} \mid \mathbf{P}) & \text{if } a \geq \frac{1}{1+k} \end{cases}$$

As $U_A(a, s_R^*; \mathbf{P})$ is strictly decreasing in a on $\left[0, \frac{1}{1+k}\right]$ and strictly increasing in a on $\left[\frac{1}{1+k}, 1\right]$ if and only if

$$(1 + k)\, \mu(\overline{\theta} \mid \mathbf{P}) > 2$$

the only candidates for best responses are the boundary solutions zero or one, respectively. Because of

$$\begin{aligned} U_A(0, s_R^*; \mathbf{P}) &\leq U_A(1, s_R^*; \mathbf{P}) \\ &\Leftrightarrow \\ 1 &\leq \frac{k}{2} \mu(\overline{\theta} \mid \mathbf{P}) \end{aligned} \tag{9}$$

we obtain the following result.

Proposition 3. *Fix the receiver's SPNE strategy s_R^*. Depending on the return parameter $k \geq 1$ the best responses of the risk-neutral agent with identity $\mathbf{P}$ are given as follows*

$$f_A(\mathbf{P}) = \begin{cases} 0 & \text{if } k\mu(\overline{\theta} \mid \mathbf{P}) < 2 \\ \{0, 1\} & \text{if } k\mu(\overline{\theta} \mid \mathbf{P}) = 2 \\ 1 & \text{if } k\mu(\overline{\theta} \mid \mathbf{P}) > 2 \end{cases}$$

To deal with the (non-generic) case of indifference, we restrict attention to the agent's SPNE strategy s_A^* such that

$$s_A^*(\mathbf{P}) = \begin{cases} 0 & \text{if } k\mu(\overline{\theta} \mid \mathbf{P}) \leq 2 \\ 1 & \text{if } k\mu(\overline{\theta} \mid \mathbf{P}) > 2 \end{cases}$$

The agent with the principal's identity will send the maximal amount to the receiver if the expected return $k\mu(\overline{\theta} \mid \mathbf{P})$ is sufficiently high. Else the agent will not send anything. In particular, the receiver will never receive any money from this agent if $k < 2$ regardless of how large the probability of cooperation might be.

3.3 Principal

Having solved for the SPNE strategies s_A^*, s_R^* of the agent and the receiver, respectively, we can move up to the first stage of the game. The principal strictly prefers in an SPNE the agent with the receiver's identity $\mathbf{R}$ over the agent with her own identity $\mathbf{P}$ if and and only if

$$U_P(\mathbf{R}, s_A^*, s_R^*) > U_P(\mathbf{P}, s_A^*, s_R^*)$$

where

$$U_P(\mathbf{R}, s_A^*, s_R^*) = \sum_{\theta \in \{\underline{\theta}, \overline{\theta}\}} (1 - s_A^*(\mathbf{R}) + s_R^*(s_A^*(\mathbf{R}), \theta) k s_A^*(\mathbf{R})) \mu(\theta \mid \mathbf{R}) \tag{10}$$

and

$$U_P(\mathbf{P}, s_A^*, s_R^*) = \sum_{\theta \in \{\underline{\theta}, \overline{\theta}\}} (1 - s_A^*(\mathbf{P}) + s_R^*(s_A^*(\mathbf{P}), \theta) k s_A^*(\mathbf{P})) \mu(\theta \mid \mathbf{P}) \tag{11}$$

Substituting the SPNE strategies s_A^*, s_R^* from the previous analysis in (10) and (11), respectively, results in

$$U_p(\mathbf{R}, s_A^*, s_R^*) = \frac{k}{2} \mu\left(\overline{\theta} \mid \mathbf{R}\right)$$

and

$$U_P(\mathbf{P}, s_A^*, s_R^*) = \begin{cases} 1 & \text{if } k\mu\left(\overline{\theta} \mid \mathbf{P}\right) \leq 2 \\ \frac{k}{2}\mu\left(\overline{\theta} \mid \mathbf{P}\right) & \text{if } k\mu\left(\overline{\theta} \mid \mathbf{P}\right) \geq 2 \end{cases}$$

If $k\mu\left(\overline{\theta} \mid \mathbf{P}\right) \leq 2$, the principal strictly prefers the agent with identity $\mathbf{R}$ over the agent with identity $\mathbf{P}$ if and only if

$$\frac{k}{2}\mu\left(\overline{\theta} \mid \mathbf{R}\right) > 1$$

If we have instead $k\mu\left(\overline{\theta} \mid \mathbf{P}\right) \geq 2$, the principal strictly prefers the agent with identity $\mathbf{R}$ over the agent with identity $\mathbf{P}$ if and only if

$$\frac{k}{2}\mu\left(\overline{\theta} \mid \mathbf{R}\right) > \frac{k}{2}\mu\left(\overline{\theta} \mid \mathbf{P}\right)$$

Let us summarize these findings.

Theorem 1.

(i) *Suppose that the model's parameters satisfy*

$$k\mu\left(\overline{\theta} \mid \mathbf{P}\right) \leq 2$$

Then the principal strictly prefers in an SPNE the agent with the receiver's identity $\mathbf{R}$ *over the agent with her own identity* $\mathbf{P}$ *if and only if*

$$k\mu\left(\overline{\theta} \mid \mathbf{R}\right) > 2$$

(ii) *Suppose that the model's parameters satisfy instead*

$$k\mu\left(\overline{\theta} \mid \mathbf{P}\right) \geq 2 \tag{12}$$

Then the principal strictly prefers in an SPNE the agent with the receiver's identity $\mathbf{R}$ *over the agent with her own identity* $\mathbf{P}$ *if and only if*

$$\mu\left(\overline{\theta} \mid \mathbf{P}\right) < \mu\left(\overline{\theta} \mid \mathbf{R}\right)$$

Observe that Theorem 1 implies that the principal strictly prefers the agent with the receiver's identity if and only if (i) the expected return $k\mu\left(\overline{\theta} \mid \mathbf{R}\right)$ is strictly above the model-specific threshold value of two and (ii) the likelihood of cooperation is strictly greater for the agent with the receiver's than with the principal's identity. The fact that the expected material payoff maximizing principal of our model would never choose the agent with the receiver's identity if

$$k\mu\left(\overline{\theta} \mid \mathbf{R}\right) \leq 2 \tag{13}$$

is directly driven by our choice of the Fehr and Schmidt (1999) utility specification (4) which implies that the receiver will share half of the total return on a maximal investment with the principal. If (13) holds, the zero-investment, which would be chosen by the agent with the principal's identity, is therefore optimal from the principal's perspective.

4 Introducing a risk-averse agent

So far we had assumed that the agent with the principal's identity $\mathbf{P}$ has exactly the same preferences as the principal. This section drives a wedge between the preferences of the risk-neutral principal and the loyal agent, who cannot observe the receiver's cooperation type, by assuming that the agent is risk-averse. To distinguish between a risk-neutral principal and a risk-averse agent has a long tradition in the classical principal-agent literature. This difference in risk-preferences for a given principal-agent relationship is typically justified by the assumption that the principal is, in contrast to the single-individual agent, a large organization that exploits the law of large numbers by operating many agents.[10] This section follows this standard assumption.

[10]To quote from the standard textbook on contract theory by Bolton and Dewatripont (2005, p.68): "Because of imperfect insurance markets, moreover, parties to the contract are risk averse. This observation is especially true of workers, whose human capital is much less easily diversifiable than financial capital."

4.1 Agent with identity P

We consider a risk-averse agent who is an expected utility maximizer with a strictly concave Bernoulli utility function $u : [0,1] \to \mathbb{R} \cup \{-\infty\}$ which is strictly increasing and continuously differentiable on $(0,1]$. We call this agent *risk-averse** if u becomes infinitely steep whenever x converges to zero, i.e., $\lim_{x\to 0} u'(x) = \infty$. Risk-averse* Bernoulli utility functions are, for example, CRRA (*constant relative risk aversion*) utility functions.

Fix the SPNE strategy s_R^* of the receiver. The expected utility of the risk-averse agent with identity $\mathbf{P}$ from action $a \in [0,1]$ is then given as

$$\begin{aligned} U_A\left(a, s_R^*; \mathbf{P}\right) &= \sum_{\theta \in \left\{\underline{\theta}, \overline{\theta}\right\}} u\left(1 - a + s_R^*\left(\theta, a\right) ka\right) \mu\left(\theta \mid \mathbf{P}\right) \\ &= u\left(1-a\right)\left(1 - \mu\left(\overline{\theta} \mid \mathbf{P}\right)\right) + u\left(1 - a + s_R^*\left(\overline{\theta}, a\right) ka\right) \mu\left(\overline{\theta} \mid \mathbf{P}\right) \end{aligned} \tag{14}$$

We derive in the Appendix the following proposition.

Proposition 4. *Fix the receiver's SPNE strategy* s_R^**.*

(a) *If* $k = 1$*, the unique best response, and therefore the unique SPNE action, of the (strictly) risk-averse agent with identity* $\mathbf{P}$ *is given as follows*

$$s_A^*\left(\mathbf{P}\right) = f_A\left(\mathbf{P}\right) = 0$$

(b) *If* $k > 1$*, there are exactly two candidates for a best response of the risk-averse* agent: either* $f_A\left(\mathbf{P}\right) = 0$ *or* $f_A\left(\mathbf{P}\right) = \hat{a}$ *such that* $\hat{a} \in \left(\frac{1}{1+k}, 1\right)$ *and* $\hat{a}$ *is pinned down by the following first order condition*

$$\frac{u'\left(\frac{1}{2} + \frac{k-1}{2}\hat{a}\right)}{u'\left(1-\hat{a}\right)} = \frac{2}{k-1} \frac{\left(1 - \mu\left(\overline{\theta} \mid \mathbf{P}\right)\right)}{\mu\left(\overline{\theta} \mid \mathbf{P}\right)} \tag{15}$$

The important insight of Proposition 4 is that $\hat{a}$ is always bounded away from one for arbitrary values of $k \geq 1$. For $k > 1$ this property is guaranteed by the assumption that the concave u becomes infinitely steep whenever the payoffs converge to zero. This can be directly seen from the first-order condition (15): if $\hat{a} \to 1$, the equation (15) would be violated as we have a strictly positive number on the RHS whereas the LHS converges towards zero because of $\lim_{\hat{a}\to 1} u'\left(1-\hat{a}\right) = \infty$.

In words: it can never be optimal for the risk-averse* agent to send the maximal amount of one to the receiver because this agent fears too much the possibility that the receiver turns out to be of the non-cooperative type. As a consequence, we can restrict in the subsequent analysis (i.e., Theorem 2) attention to either the interior solution pinned down by the first-order condition (15) or the boundary solution where the agent does not send any money to the receiver.

Remark. The risk-averse* agent of Proposition 4(b) chooses as SPNE action $s_A^*(\mathbf{P}) = \hat{a}$ over $s_A^*(\mathbf{P}) = 0$ if and only if

$$U_A(0, s_R^*; \mathbf{P}) \leq U_A(\hat{a}, s_R^*; \mathbf{P}) \tag{16}$$

Proposition 4(b) remains silent about any parameter conditions that would establish inequality (16) for general Bernoulli utility functions. We will come back to inequality (16) in Section 4.3 when we consider the special case of a logarithmic Bernoulli utility function.

4.2 Principal

Suppose that the principal has chosen the agent with identity $\mathbf{P}$ who sends amount $s_A(\mathbf{P}) = a$ to the receiver. Given the receiver's SPNE strategy s_R^* the expected utility of the risk-neutral principal is then identical to the expected utility (8) of the risk-neutral agent with identity $\mathbf{P}$, i.e.,

$$U_P(\mathbf{P}, s_A(\mathbf{P}) = a, s_R^*) = \begin{cases} 1 - a & \text{if } a \leq \frac{1}{1+k} \\ 1 - a + \frac{(1+k)a-1}{2}\mu\left(\overline{\theta} \mid \mathbf{P}\right) & \text{if } a \geq \frac{1}{1+k} \end{cases} \tag{17}$$

By the same argument (9) as for the risk-neutral agent, the risk-neutral principal wants this agent to send either zero, i.e., $s_A(\mathbf{P}) = 0$, or the maximal amount of one, i.e., $s_A(\mathbf{P}) = 1$, whereby

$$\begin{array}{rcl} U_P(\mathbf{P}, s_A(\mathbf{P}) = 0, s_R^*) & \leq & U_P(\mathbf{P}, s_A(\mathbf{P}) = 1, s_R^*) \\ & \Leftrightarrow & \\ 1 & \leq & \frac{1}{2}k\mu\left(\overline{\theta} \mid \mathbf{P}\right) \end{array}$$

which is equivalent to parameter condition (12). In contrast to the risk-neutral agent, however, the risk-averse* agent will either choose $s_A(\mathbf{P}) = 0$ or $s_A(\mathbf{P}) = \hat{a} \in \left(\frac{1}{1+k}, 1\right)$ as an SPNE action. Let us assume that this agent chooses $\hat{a}$. Under the parameter condition (12), we obtain

$$\begin{array}{rcl} U_p(\mathbf{P}, s_A^*(\mathbf{P}) = \hat{a}, s_R^*) & < & U_p(\mathbf{P}, s_A^*(\mathbf{P}) = 1, s_R^*) \\ & = & \left(\frac{1}{2}k\right)\mu\left(\overline{\theta} \mid \mathbf{P}\right) \end{array}$$

By continuity of the principal's expected utility function (17) in $\mu\left(\overline{\theta} \mid \mathbf{P}\right)$, there exists some sufficiently small $\varepsilon > 0$ such that

$$\mu\left(\overline{\theta} \mid \mathbf{R}\right) = \mu\left(\overline{\theta} \mid \mathbf{P}\right) - \varepsilon \tag{18}$$

and

$$\begin{aligned} U_p\left(\mathbf{P}, s_A^*\left(\mathbf{P}\right) = 0, s_R^*\right) &< U_p\left(\mathbf{P}, s_A^*\left(\mathbf{P}\right) = \hat{a}, s_R^*\right) \\ &< \left(\frac{1}{2}k\right)\mu\left(\overline{\theta} \mid \mathbf{R}\right) \\ &= U_p\left(\mathbf{R}, s_A^*\left(\mathbf{R}\right) = 1, s_R^*\right) \end{aligned}$$

That is, whenever the agent of Proposition 4(b) chooses $\hat{a}$ we can always find some $\varepsilon > 0$ in (18) such that the principal strictly prefers the agent with identity $\mathbf{R}$. To be precise about the possible values of ε, observe that

$$\begin{aligned} U_p\left(\mathbf{P}, s_A^*\left(\mathbf{P}\right) = \hat{a}, s_R^*\right) &\leq U_p\left(\mathbf{R}, s_A^*\left(\mathbf{R}\right) = 1, s_R^*\right) \\ &\Leftrightarrow \\ 1 - \hat{a} + \frac{(1+k)\,\hat{a} - 1}{2}\mu\left(\overline{\theta} \mid \mathbf{P}\right) &\leq \left(\frac{1}{2}k\right)\left(\mu\left(\overline{\theta} \mid \mathbf{P}\right) - \varepsilon\right) \\ &\Leftrightarrow \\ \varepsilon &\leq \frac{(1+k)\,\mu\left(\overline{\theta} \mid \mathbf{P}\right) - 2}{k}\left(1 - \hat{a}\right) \end{aligned}$$

Let us summarize the above analysis.

Theorem 2. *Suppose that the model's parameters satisfy*

$$k\mu\left(\overline{\theta} \mid \mathbf{P}\right) \geq 2 \tag{19}$$

If the risk-averse agent with identity* $\mathbf{P}$ *chooses the SPNE action* $s_A\left(\mathbf{P}\right) = \hat{a}$ *of Proposition 4(b), the principal strictly prefers in an SPNE the agent with identity* $\mathbf{R}$ *if and only if*

$$\mu\left(\overline{\theta} \mid \mathbf{R}\right) > \mu\left(\overline{\theta} \mid \mathbf{P}\right) - \varepsilon^*$$

such that

$$\varepsilon^* = \frac{(1+k)\,\mu\left(\overline{\theta} \mid \mathbf{P}\right) - 2}{k}\left(1 - \hat{a}\right)$$

Observe that ε^* is always strictly positive. In situations where the loyal but risk-averse* agent chooses $\hat{a}$ over zero, the principal thus strictly prefers the agent with the receiver's identity $\mathbf{R}$ even if the cooperation probability $\mu\left(\overline{\theta} \mid \mathbf{R}\right)$ is (slightly) smaller than $\mu\left(\overline{\theta} \mid \mathbf{P}\right)$. This implication of Theorem 2 for the risk-averse* agent is in contrast to Theorem 1 where parameter condition (19) implied that the principal will choose the agent with identity $\mathbf{R}$ over the risk-neutral agent with identity $\mathbf{P}$ if and only if the probability $\mu\left(\overline{\theta} \mid \mathbf{R}\right)$ is strictly greater than the probability $\mu\left(\overline{\theta} \mid \mathbf{P}\right)$.

4.3 Illustrative example: Logarithmic Bernoulli utility

To illustrate Proposition 4 and Theorem 2 through a closed-form example, let us consider the logarithmic Bernoulli utility function which is of the CRRA form with relative risk aversion parameter one:

$$u(x) = \begin{cases} -\infty & \text{if } x = 0 \\ \ln(x) & \text{else} \end{cases}$$

The following result is proved in the Appendix.

Proposition 5. *Suppose that the parameter condition (19) is satisfied. If the agent with identity* $\mathbf{P}$ *has a logarithmic Bernoulli utility function, its unique best response, and therefore its unique SPNE action, is given as follows*

$$s_A^*(\mathbf{P}) = f_A(\mathbf{P}) = \frac{k\mu\left(\overline{\theta} \mid \mathbf{P}\right) - 1}{k-1}$$

Moreover, $\varepsilon^* > 0$ *of Theorem 2 becomes*

$$\varepsilon^* = \frac{(1+k)\,\mu\left(\overline{\theta} \mid \mathbf{P}\right) - 2}{k-1}\left(1 - \mu\left(\overline{\theta} \mid \mathbf{P}\right)\right) \tag{20}$$

To give a concrete example, suppose that $k = 4$ and $\mu\left(\overline{\theta} \mid \mathbf{P}\right) = \frac{1}{2}$. For these parameter values condition (19) holds with equality so that the risk-neutral principal would like the agent with her own identity to either make a zero-investment or to invest the maximal amount of one whereby the principal is exactly indifferent between both actions. The risk-averse* agent with identity $\mathbf{P}$, however, will choose as SPNE action

$$s_A^*(\mathbf{P}) = f_A(\mathbf{P}) = \frac{1}{3}$$

which is strictly suboptimal from the principal's perspective. Note that (20) becomes $\varepsilon^* = \frac{1}{12}$ so that the principal strictly prefers an agent with the receiver's identity $\mathbf{R}$ if and only if

$$\mu\left(\overline{\theta} \mid \mathbf{R}\right) > \frac{1}{2} - \frac{1}{12} = \frac{5}{12}$$

or, equivalently, if and only if the expected return from this choice satisfies

$$k\mu\left(\overline{\theta}\mid\mathbf{R}\right)>4\cdot\frac{5}{12}=\frac{5}{3} \tag{21}$$

In words: the principal will choose the agent with the receiver's identity even if this agent only brings about a 41.7% chance of cooperation compared to the 50% chance of cooperation that would result from choosing the loyal agent. Moreover, by (21), the principal would now choose the agent with the receiver's identity even in situations where the expected return $k\mu\left(\overline{\theta}\mid\mathbf{R}\right)$ falls below the threshold value of two (cf. Theorem 1) as long as the expected return is above 1.667.

5 Concluding remarks

Members of a given social group often exhibit an in-group bias in their preferences according to which they favour members of their own group identity over people with different group-identities. We have constructed a trust game with delegated decision making to study the impact of such in-group bias in situations where a principal has to choose the group identity of an agent who deals with a receiver on the principal's behalf . By assumption, the principal and the receiver have different group identities so that the principal has to choose either an agent with her own or with the receiver's group identity. Given that the agent of our model exhibits a strong in-group bias, the answer to the principal's choice problem seems to be straightforward: The principal is going to choose a loyal agent of her own group identity over a disloyal agent of the receiver's group identity.

To make things interesting–and to make our model more relevant for real-life choice situations–we have introduced two additional layers of complexity. In a first scenario, we assume that an agent with the receiver's group identity might increase the likelihood that the receiver will cooperate towards the principal's objectives. In a second scenario, we additionally consider the possibility that the principal and the agent with the principal's identity have misaligned preferences in the sense that the principal is risk-neutral whereas the agent is risk-averse. With these two added layers of complexity we have addressed the following research question:

> Under which conditions on the model's structure and payoff parameters will the principal choose the agent of her own identity over the agent with the receiver's identity and vice versa?

To answer this question, we have solved for the principal's choice of the agent's group identity as the outcome of a subgame-perfect Nash equilibrium. All our analytical results have been sharp with respect to the model's parameter values. Let us summarize the qualitative main insights from our analysis. Under suitable parameter conditions, the principal chooses the agent of the receiver's group identity over the loyal agent of her own identity if and only if this choice increases the likelihood of the receiver being cooperative. But even without such increase in the receiver's willingness to cooperate, the risk-neutral principal might strictly prefer the agent of the receiver's identity whenever the agent of her own identity is too risk-averse from her perspective.

Appendix: Formal proofs

Proof of Proposition 1. If $a = 0$, we trivially have that $\lambda^* \in [0,1]$, i.e., any type of agent is indifferent between all possible actions if he cannot send back any money anyway.

Suppose now that $a > 0$. The utility specification (4) implies that there cannot be a maximizer $\lambda^* > 0$ such that $\pi_R(a,\lambda^*) < \pi_P(a,\lambda^*)$. Focus therefore on the relevant case

$$\begin{aligned} \pi_R(a,\lambda^*) &\geq \pi_P(a,\lambda^*) \\ &\Leftrightarrow \\ \frac{(1+k)a-1}{2ka} &\geq \lambda^* \end{aligned}$$

implying that any maximizer $\lambda^* \geq 0$ must lie in the interval

$$\left[0, \max\left\{0, \frac{(1+k)a-1}{2ka}\right\}\right] \tag{22}$$

Taking the first-order derivative of

$$\begin{aligned} U_R(a,\lambda;\theta) &= \pi_R(a,\lambda) - \theta(\pi_R(a,\lambda) - \pi_P(a,\lambda)) \\ &= (1-\theta)(1-\lambda)ka + \theta(1-a+\lambda ka) \end{aligned}$$

with respect to λ

$$\begin{aligned} \frac{d}{d\lambda}U_R(a,\lambda;\theta) &\geq 0 \\ &\Leftrightarrow \\ -(1-\theta)ka + \theta(ka) &\geq 0 \end{aligned}$$

shows that

$$\begin{aligned} \frac{d}{d\lambda}U_R(a,\lambda;\theta) &= 0 \text{ if and only if } a = 0 \text{ or } \theta = \frac{1}{2} \\ \frac{d}{d\lambda}U_R(a,\lambda;\theta) &> 0 \text{ if and only if } a > 0 \text{ and } \theta > \frac{1}{2} \\ \frac{d}{d\lambda}U_R(a,\lambda;\theta) &< 0 \text{ if and only if } a > 0 \text{ and } \theta < \frac{1}{2} \end{aligned}$$

For $\underline{\theta} \in \left[0, \frac{1}{2}\right)$ the maximizing λ^* must coincide with the lower boundary in (22). For $\overline{\theta} \in \left(\frac{1}{2}, 1\right]$ the maximizing λ^* must coincide with the upper boundary in (22). Collecting the above results gives us Proposition 1.

□□

Proof of Proposition 2. Part (a). Consider at first the non-cooperative type $\underline{\theta}$. Since $s_R^*(a, \underline{\theta}) = 0$, (7) becomes

$$\begin{aligned} U_A(a, 0; \mathbf{R}, \underline{\theta}) &= (1 - \underline{\theta}) ka + \underline{\theta}(1 - a) \\ &= (k - (k+1)\underline{\theta}) a + \underline{\theta} \end{aligned} \tag{23}$$

Consequently, the agent's utility (23) is strictly increasing in a if and only if

$$\begin{aligned} k - (k+1)\underline{\theta} &> 0 \\ &\Leftrightarrow \\ \frac{k}{k+1} &> \underline{\theta} \end{aligned}$$

which is always satisfied under our parameter assumptions $k \geq 1$ and $\underline{\theta} < \frac{1}{2}$. This proves part (a) of the proposition.

Part (b). Turn now to the cooperative type $\overline{\theta}$. Recall from the proof of Proposition 1 that the receiver's payoff is weakly greater than the principal's payoff if and only if

$$\frac{(1+k)a - 1}{2ka} \geq s_R^*\left(a, \overline{\theta}\right)$$

In what follows we compare the two possible cases $s_R^*\left(a, \overline{\theta}\right) = 0$ and $s_R^*\left(a, \overline{\theta}\right) = \frac{(1+k)a-1}{2ka}$, respectively.

Case (i). Suppose that $s_R^*\left(a, \overline{\theta}\right) = 0$. Then any utility maximizing a^* must thus satisfy the lower boundary condition

$$a^* \geq \frac{1}{1+k} \tag{24}$$

Substituting $s_R^*\left(a, \overline{\theta}\right) = 0$ in (7) gives

$$U_A\left(a, 0; \mathbf{R}, \overline{\theta}\right) = \left(k - (k+1)\overline{\theta}\right) a + \overline{\theta}$$

which is strictly increasing in a if and only if

$$\begin{aligned} k - (k+1)\overline{\theta} &> 0 \\ &\Leftrightarrow \\ \frac{k}{k+1} &> \overline{\theta} \end{aligned} \tag{25}$$

Note that (25) together with the boundary condition (24) implies

$$a^* = \begin{cases} 1 & \text{if } \frac{k}{k+1} > \overline{\theta} \\ \left[\frac{1}{1+k}, 1\right] & \text{if } \frac{k}{k+1} = \overline{\theta} \\ \frac{1}{1+k} & \text{if } \frac{k}{k+1} < \overline{\theta} \end{cases}$$

By Proposition 1, $s_R^*\left(a^*,\overline{\theta}\right)=0$ requires

$$a^* \leq \frac{1}{1+k}$$

Consequently, $s_R^*\left(a^*,\overline{\theta}\right)=0$ together with

$$a^* = \frac{1}{1+k}$$

are candidate SPNE actions if and only if the model parameters satisfy

$$\frac{k}{k+1} \leq \overline{\theta}. \tag{26}$$

The corresponding utility for case (i) is

$$U_A\left(a^*=\frac{1}{1+k}, s_R^*\left(a^*,\overline{\theta}\right)=0; \mathbf{R},\overline{\theta}\right)=\frac{k}{k+1} \tag{27}$$

Case (ii). Suppose now that

$$\begin{aligned} s_R^*\left(a,\overline{\theta}\right) &= \frac{(1+k)\,a-1}{2ka} > 0 \\ &\Leftrightarrow \\ a &> \frac{1}{1+k} \end{aligned}$$

Substitution in (7) gives after some transformations

$$\begin{aligned} & U_A\left(a, s_R^*\left(a,\overline{\theta}\right); \mathbf{R},\overline{\theta}\right) \qquad (28) \\ = & \left(1-\overline{\theta}\right)\left(1-\frac{(1+k)\,a-1}{2ka}\right)ka+\overline{\theta}\left(1-a+\frac{(1+k)\,a-1}{2ka}ka\right) \\ = & \frac{(k-1)\,a+1}{2} \end{aligned}$$

If $k>1$, the utility (28) is strictly increasing in a so that $a^{**}=1$ is the unique maximizer of (28). The corresponding utility for case (ii) is

$$U_A\left(a^{**}=1, s_R^*\left(a^{**},\overline{\theta}\right)=\frac{1}{2}; \mathbf{R},\overline{\theta}\right)=\frac{k}{2}$$

Comparing the utilities of both candidate solutions, (27) and (28) respectively, shows that

$$\begin{aligned} U_A\left(a^{**}=1, s_R^*\left(a^{**},\overline{\theta}\right)=\frac{1}{2}; \mathbf{R},\overline{\theta}\right) &> U_A\left(a^*=\frac{1}{1+k}, s_R^*\left(a^*,\overline{\theta}\right)=0; \mathbf{R},\overline{\theta}\right) \\ &\Leftrightarrow \\ k &> 1 \end{aligned}$$

Thus, for $k > 1$ the unique SPNE action of this agent will be $a^{**} = 1$. If $k = 1$, the utility (28) is constantly

$$U_A\left(a, s_R^*\left(a, \overline{\theta}\right); \mathbf{R}, \overline{\theta}\right) = \frac{1}{2}$$

for all $a > \frac{1}{2}$. Similarly, the utility (27) becomes

$$U_A\left(a, s_R^*\left(a, \overline{\theta}\right) = 0; \mathbf{R}, \overline{\theta}\right) = \frac{1}{2}$$

for $a = \frac{1}{2}$ whereby condition (26) is satisfied. This shows that any $a^* \in \left[\frac{1}{2}, 1\right]$ is a best response given $s_R^*\left(a, \overline{\theta}\right)$ for the special case $k = 1$.

□□

Proof of Proposition 4. Part (a). For $k = 1$ we have

$$s_R^*\left(a, \overline{\theta}\right) = \begin{cases} 0 & \text{if } a \leq \frac{1}{2} \\ \frac{2a-1}{2a} & \text{if } a \geq \frac{1}{2} \end{cases}$$

Case (i). Suppose that $a \leq \frac{1}{2}$ so that (14) becomes

$$U_A\left(a, s_R^*; \mathbf{P}\right) = u\left(1 - a\right)$$

which is maximized at $a^{**} = 0$. That is, $a^{**} = 0$ is, as the unique local maximizer on the interval $\left[0, \frac{1}{2}\right]$, one candidate for the global maximizer on $[0, 1]$.

Case (ii). Suppose now that $a \geq \frac{1}{2}$. Substitution in (14) gives

$$U_A\left(a, s_R^*; \mathbf{P}\right) = u\left(1 - a\right)\left(1 - \mu\left(\overline{\theta} \mid \mathbf{P}\right)\right) + u\left(\frac{1}{2}\right)\mu\left(\overline{\theta} \mid \mathbf{P}\right)$$

which strictly decreases in a. Consequently, $a^* = \frac{1}{2}$ is the local maximizer on the interval $\left[\frac{1}{2}, 1\right]$. As we had already established under Case (i) that

$$U_A\left(0, s_R^*; \mathbf{P}\right) > U_A\left(\frac{1}{2}, s_R^*; \mathbf{P}\right)$$

$a^{**} = 0$ is the global maximizer on $[0, 1]$.

Part (b). For $k > 1$ we have

$$s_R^*\left(a, \overline{\theta}\right) = \begin{cases} 0 & \text{if } a \leq \frac{1}{1+k} \\ \frac{(1+k)a-1}{2ka} & \text{if } a \geq \frac{1}{1+k} \end{cases}$$

If $a \leq \frac{1}{1+k}$, we trivially obtain

$$U_A\left(a, s_R^*; \mathbf{P}\right) = u\left(1 - a\right)$$

so that $a^{**}=0$ is the unique local maximizer on the interval $\left[0,\frac{1}{1+k}\right]$. By continuity of $U_A\left(a,s_R^*;\mathbf{P}\right)$, the only other candidate for a global maximizer would be the local maximizer, denoted a^*, on the interval $\left(\frac{1}{1+k},1\right]$ (provided that it exists).

Suppose therefore that $a>\frac{1}{1+k}$. Substitution in (14) gives after some transformation

$$U_A\left(a,s_R^*;\mathbf{P}\right)=u\left(1-a\right)\left(1-\mu\left(\overline{\theta}\mid\mathbf{P}\right)\right)+u\left(\frac{1}{2}+\frac{k-1}{2}a\right)\mu\left(\overline{\theta}\mid\mathbf{P}\right)$$

Taking the first order derivative

$$-u'\left(1-a\right)\left(1-\mu\left(\overline{\theta}\mid\mathbf{P}\right)\right)+\frac{k-1}{2}u'\left(\frac{1}{2}+\frac{k-1}{2}a\right)\mu\left(\overline{\theta}\mid\mathbf{P}\right)\leq 0$$

$$\Leftrightarrow$$

$$\frac{u'\left(\frac{1}{2}+\frac{k-1}{2}a\right)}{u'\left(1-a\right)}\leq\frac{2}{k-1}\frac{\left(1-\mu\left(\overline{\theta}\mid\mathbf{P}\right)\right)}{\mu\left(\overline{\theta}\mid\mathbf{P}\right)}$$

shows that $U_A\left(a,s_R^*;\mathbf{P}\right)$ is either strictly decreasing in a because of $\lim_{a\to1}u'\left(\frac{1}{2}+\frac{k-1}{2}a\right)<\infty$ together with our assumption $\lim_{a\to1}u'\left(1-a\right)=\infty$, or we have a unique maximizer $\hat{a}$ on $(-\infty,1)$ pinned down by the first order condition

$$\frac{u'\left(\frac{1}{2}+\frac{k-1}{2}\hat{a}\right)}{u'\left(1-\hat{a}\right)}=\frac{2}{k-1}\frac{\left(1-\mu\left(\overline{\theta}\mid\mathbf{P}\right)\right)}{\mu\left(\overline{\theta}\mid\mathbf{P}\right)}\tag{29}$$

Consequently, $a^*=\hat{a}<1$ is the unique local maximizer on the interval $\left(\frac{1}{1+k},1\right]$ if and only if $\hat{a}>\frac{1}{1+k}$.

To summarize for $k>1$. If the global maximizer is unique on $[0,1]$, it is either zero or $\hat{a}>\frac{1}{1+k}$ pinned down by (29). For the non-generic case that

$$U_A\left(0,s_R^*;\mathbf{P}\right)=U_A\left(\hat{a},s_R^*;\mathbf{P}\right)$$

the set of global maximizers on $[0,1]$ is non-convex and consists exactly of zero and $\hat{a}$.
□□

Proof of Proposition 5. The first order condition (15) becomes for a logarithmic Bernoulli utility function

$$\frac{k-1}{2}\frac{\left(1-\hat{a}\right)}{\frac{1}{2}+\frac{k-1}{2}\hat{a}}=\frac{\left(1-\mu\left(\overline{\theta}\mid\mathbf{P}\right)\right)}{\mu\left(\overline{\theta}\mid\mathbf{P}\right)}$$

$$\Leftrightarrow$$

$$\hat{a}=\frac{k\mu\left(\overline{\theta}\mid\mathbf{P}\right)-1}{k-1}$$

First, note that $\hat{a}$ satisfies the necessary boundary condition

$$\hat{a} > \frac{1}{1+k}$$

if and only if

$$\begin{aligned} \frac{k\mu\left(\overline{\theta} \mid \mathbf{P}\right) - 1}{k-1} &> \frac{1}{1+k} \\ &\Leftrightarrow \\ k\mu\left(\overline{\theta} \mid \mathbf{P}\right) &> 2 - \mu\left(\overline{\theta} \mid \mathbf{P}\right) \end{aligned} \tag{30}$$

Next observe that the agent strictly prefers the non-zero amount $\hat{a} > \frac{1}{1+k}$ to zero if and only if

$$U_A\left(\hat{a}, s_R^*; \mathbf{P}\right) > U_A\left(0, s_R^*; \mathbf{P}\right) \tag{31}$$

We have

$$U_A\left(0, s_R^*; \mathbf{P}\right) = \ln(1) = 0$$

and

$$U_A\left(\hat{a}, s_R^*; \mathbf{P}\right) = \ln\left(\frac{k}{k-1}\left(1 - \mu\left(\overline{\theta} \mid \mathbf{P}\right)\right)\right)\left(1 - \mu\left(\overline{\theta} \mid \mathbf{P}\right)\right) + \ln\left(\frac{1}{2}k\mu\left(\overline{\theta} \mid \mathbf{P}\right)\right)\mu\left(\overline{\theta} \mid \mathbf{P}\right)$$

so that inequality (31) becomes

$$\ln\left(\frac{k}{k-1}\left(1 - \mu\left(\overline{\theta} \mid \mathbf{P}\right)\right)\right) > \ln\left(\frac{2}{k-1}\frac{\left(1 - \mu\left(\overline{\theta} \mid \mathbf{P}\right)\right)}{\mu\left(\overline{\theta} \mid \mathbf{P}\right)}\right)\mu\left(\overline{\theta} \mid \mathbf{P}\right) \tag{32}$$

To interpret (32), observe that

$$\begin{aligned} \frac{k}{k-1}\left(1 - \mu\left(\overline{\theta} \mid \mathbf{P}\right)\right) &\geq \frac{2}{k-1}\frac{\left(1 - \mu\left(\overline{\theta} \mid \mathbf{P}\right)\right)}{\mu\left(\overline{\theta} \mid \mathbf{P}\right)} \\ &\Leftrightarrow \\ k\mu\left(\overline{\theta} \mid \mathbf{P}\right) &\geq 2 \end{aligned}$$

Consequently, (32) holds for any given value $\mu\left(\overline{\theta} \mid \mathbf{P}\right) < 1$ whenever the parameter condition (19) is satisfied. Finally note that the boundary condition (30) is automatically satisfied whenever (19) holds.

In words: If the payoff parameter condition (19) of Theorem 2 is satisfied, the agent with logarithmic Bernoulli utility function will choose the investment amount $\hat{a}$ of Proposition 4(b) over the zero-investment. Collecting the above results gives us Proposition 5.

□□

References

Ashraf, N., Bohnet, I., & Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, 9, 193-208.

Balliet, D., Van Lange, P. A. (2013). Trust, punishment, and cooperation across 18 societies: a meta-analysis. *Perspectives on Psychological Science*, 8, 363-379.

Balliet, D., Wu, J., & De Dreu, C.K. (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological Bulletin*, 140, 1556-1581.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10, 122-142.

Bolton, P., Dewatripont, M. (2005). *Contract theory.* The MIT Press.

Bolton, G. E., Ockenfels, A. (1986). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90, 166-193.

Bornhorst, F., Ichino, A., Kirchkamp, O., Schlag, K. H., & Winter, E. (2010). Similarities and differences when building trust: the role of cultures. *Experimental Economics*, 13, 260-283.

Burns, J. (2006). Racial stereotypes, stigma and trust in post-apartheid South Africa. *Economic Modelling*, 23, 805-821.

Chen, Y., Li, S. X. (2009). Group identity and social preference. *American Economic Review*, 99, 431–457.

Chowdhury, S., Jeon, J., Ramalingam, A. (2016). Identity and group conflict. *European Economic Review*, 90, 107-121.

Currarini, S., Mengel, F. (2016). Identity, homophily and in-group bias. *European Economic Review*, 90, 40-55.

Daskalova, V. (2018). Discrimination, social identity, and coordination: An experiment. *Games and Economic Behavior*, 107, 238-252.

Dufwenberg, M., Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, 47, 268-298.

Falk, A., Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54, 293-315.

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

Fehr, E., Gächter, S, & Kirchsteiger, G. (1996). Reciprocity as a contract enforcement device. *Econometrica*, 65, 833–860.

Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. *Quarterly Journal of Economics*, 108, 437–460.

Fershtman, C., Gneezy, U. (2001). Discrimination in a segmented society: An experimental approach. *Quarterly Journal of Economics*, 116, 351-377.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *Quarterly Journal of Economics*, 115, 811-846.

Güth, W., Levati, M. V., & Ploner, M. (2008). Social identity and trust–An experimental investigation. *The Journal of Socio-Economics*, 37, 1293-1308.

Hamann, H., Nicholls, N. (2018). Group identity in intermediated interactions: Lessons from a Trust Game with Delegation in South Africa. In D.A. Norton & A. Gunnthorsdottir (Eds) *Research in Experimental Economics*, 20, 227-264. Emerald Group Publishing Limited.

Johnson, N. D., Mislin, A. A. (2011). Trust games: a meta-analysis. *Journal of Economic Psychology*, 32, 865-889.

Morton, R. B., Ou, K., & Qin, X. (2019). Morton, R. B., K. Ou and X. Qin (2019). The effect of religion on Muslims' charitable contributions to members of a non-Muslim majority. *Journal of Public Economic Theory*, https://doi.org/10.1111/jpet.12352, to appear.

Kuwabara, K., Willer, R., Macy, M. W., Mashima, R., Terai, S., & Yamagishi, T. (2007). Culture, identity, and structure in social exchange: A web-based trust experiment in the United States and Japan. *Social Psychology Quarterly*, 70, 461-479.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, 83, 1281-1302.

Sherif, M. (2010). *The robbers cave experiment: Intergroup conflict and cooperation.* [Originally published as *Intergroup conflict and group relations*]. Wesleyan University Press.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, 223, 96-102.

Tajfel, H., Turner, J.C. (1986). The social identity theory of intergroup behavior in S. Worchel & W.G. Austin (Eds) *Psychology of Intergroup Relations.* Chicago: Nelson.

Weisel, O., Zultan, R. (2016). Social motives in intergroup conflict: group identity and perceived target of threat. *European Economic Review*, 90, 122-133.