In transparency we trust an experimental study of reputation, transparency, and signaling

Natalia Borzino [a], Enrique Fatas [b,c,*], Emmanuel Peterle [d]

[a] *Singapore ETH-Centre, Singapore*
[b] *Facultad de Ciencias Administrativas y Económicas, Universidad ICESI, Colombia*
[c] *Center for Social Norms and Behavioral Dynamics, University of Pennsylvania, USA*
[d] *CRESE EA3190 Univ. Franche-Comté F-25000, Besançon, France*

A B S T R A C T

We study transparency and reputation in a stylized investment game in which two senders interact with a receiver in a linear network with partial information. We analyze how *ex-ante* information (compatible with receiver's *reputation* building) and *ex-post* information (making the receiver's decisions *transparent*) affect behavior in the trust game. In the *Reputation* treatment, senders know the average amount returned by the receiver to senders in the previous period, but not the actions of those senders. In the *Transparency* treatment, senders are aware of the amount received by the other sender, but not the action of this other sender. Relative to a *Baseline treatment* in which senders are only informed about their own actions and outcomes, *ex-ante* and *ex-post* information disclosure have very different effects. While reputation helps some receivers, aggregate trust does not significantly increase in the *Reputation* condition. *Transparency* generates a large and significant increase in trust (amounts sent are 27% above the ones chosen in *Reputation*), being the effect largely driven by a sustained and large reduction of senders choosing not to trust at all (56% less in *Transparency* than in *Reputation* in the second half of the experiment). Our design also introduces a non-binding recommendation for both senders and receivers. Across all treatments and roles, this recommendation significantly and positively shapes individual decisions.

1. Introduction

Trust is an essential component of most social and economic interactions. Because of trust, agents are willing to productively exchange, even in situations in which standard theory predicts no exchange at all. In their recent book, Keefer and Scartascini (2022) define trust as "the belief that others will not act opportunistically. It is faith in others, in their honesty, dependability, and good will. {…} Without trust, people live in fear, not freedom; they focus on the opportunities of today rather than innovating to expand the opportunities of tomorrow" (page XVII). The benefits of trust are numerous, and go from strengthening democratic political institutions (as politicians are more likely to be held accountable by citizens, and voters will have more reasons to trust their promises), to fostering growth (as the likelihood of firms investing in risky innovations increases with the level of trust in regulatory agencies

and patent laws); organizations' performance (as employers will rely more on employees, who will be offered better working conditions), anti-crime policies (as the best police force is the one trusted by citizens, who do not fear them) or even public health (as compliance with public health guidelines during a pandemic positively and substantially depends on citizens' trust in science and government, see Bicchieri et al., 2021).

In search for a characterization of its behavioral dynamics, trust has been extensively studied by economists, psychologists and social scientists. The seminal experimental game used is the two-player investment game (Berg et al., 1995). This experimental model allows the investigation of trust and trustworthiness in a bilateral relationship between a single sender and a single receiver with similar starting positions and full information about the decisions made by the other player.[1] Berg et al. (1995), as we also do in this current study, refer to the

---

* Corresponding author.

E-mail addresses: natalia.borzino@frs.ethz.ch (N. Borzino), efatas@icesi.edu.co, efatas@upenn.edu (E. Fatas), emmanuel.peterle@univ-fcomte.fr (E. Peterle).

[1] Variants of the same game have been studied, including the systematic manipulation of resources allocated to senders and receivers, see Calabuig et al (2016) for a review.

sending decision as trust and to the returning decision as trustworthiness. The relatively simple strategic structure of the original trust game presents two important constraints. The first one is related to the number of interacting agents and the second to the amount of information available to players. In many real economic settings, including in the examples provided above, interactions are usually embedded in networks and some information about decisions in the network flows through the links.[2]

Sharing information in word-of-mouth networks and other settings may mitigate trust problems (Granovetter, 1985). Indeed, several empirical studies support the hypothesis that network embeddedness promotes trustfulness and trustworthiness in the context of full information in the network (see, for example, Buskens et al., 2010; Cassar & Rigdon, 2011; and Frey et al., 2019). These recent networked extensions of the trust game have suggested the comparative nature of trust: one's decision to trust may be affected not only by the own experience but also by the experience of others in the same network.[3] Performance information across links can also promote generalized reciprocity, as a repeated history of cooperation is spread through multiple trusted partners, which in turn builds perceptions and expectations of others (Dezső et al., 2015; Dezső et al., 2019; Dezső et al., 2022) and supports future trusting behavior (Ostrom et al., 1994; Fehr and Gatcher, 2000). In these settings, sanctions can be executed by the trustor herself or by third parties such as other trustors of the trustee.

However, the studies mentioned in the previous paragraph evaluated network embeddedness in the context of full information. This approach limits external validity as it is difficult to find real-life settings in which networks provide complete information. Our study addresses this gap studying when partial information disclosure about the receiver performance is a signal strong enough to build trust among senders and for the receiver to honor that trust. Furthermore, we study which type of partial information could most effectively boost trust and trustworthiness in networks where more than one sender interacts with the receiver.

Exploring the role and the effectiveness of partial information on trust is important for societies, organizations, and markets. Real-life examples could be found in the relationship of trust between employees with similar or different ranks or positions within organizations (e.g., employees or stakeholders interacting with a team leader or a CEO). In these settings, information exchanges in the network is generally partial and with not straightforward effects on trust. Recent research suggests that companies might benefit from raising trust within their given network structure (e.g., Austin, 2013; Brown et al., 2015).

Trust may be driven by economic and non-economic factors (Feld and Frey, 2007; Torgler, 2007), including not only performance, but how performance information is disclosed. In a survey experiment run in Argentina, Alessandro et al. (2021) shows how the link between informational policies and trust in government is mediated by government performance. The authors manipulate the information about the government's performance. If the government was presented as an overachiever player, trust in government increased, and the opposite effect was observed when presented as an underperformer, as individuals may mistrust the government because of its poor performance (Fan & Yang, 2019; Lee & Van Ryzin, 2020), or because of the lack of transparent information to interpret their accomplishments (Feld and Frey, 2007; Alm et al., 2010; Castro & Guccio, 2019).

In this paper, we study whether and to what extent different types of information shape trust when more than one sender (e.g., employees or stakeholders) interact with a receiver (e.g., team leader, CEO, or government) in an asymmetric network with two peripheral players (senders) and one central player (the receiver). As exogenously manipulating information is almost impossible in the field, we take our research to the laboratory. Previous studies support the hypothesis that full information may promote trust and trustworthiness in networked investment games. In these settings, trust is not only affected by subjects' individual experience but also by their knowledge about the experience of others.[4] Cassar and Rigdon (2011) compare settings with no information (senders observe actions and outcomes in their own investment game) with full information conditions (in which senders receive information about the amounts sent and the amounts received by others) to find that complete information increases both trust and trustworthiness. Rather than studying the effect of full information on trust and trustworthiness, we investigate whether partial information about the receiver's performance is strong enough to build trust and trustworthiness.

We study two types of partial information, one dealing with *ex-ante* information (how the receiver performed in previous rounds, *overall*) and *ex-post* information (how the receiver treated different senders' decisions) in a networked investment game with three players (two senders and one receiver). In our *Reputation treatment* (*Reputation*), we provide senders with information about the receiver decisions (the average amount returned to both senders), without revealing the individual actions of senders. In our *Transparency treatment* (*Transparency*), receivers are aware that senders will know the amounts received by both senders (again, without revealing the action of the other sender). Relative to the *Baseline treatment*, with no additional information, both conditions disclose partial information about the receiver's actions. While in *Reputation* the receiver may build a reputation on her trustworthiness before senders decide, in *Transparency* receivers are aware that senders will only receive additional information *ex-post*.

Our paper makes an additional contribution by giving participants a non-binding recommendation. As trustworthy individuals may respect socially acceptable actions, we provide participants with a rule that guarantees equal expected payoffs for senders and receivers. Both senders and receivers receive a noisy (and random) recommendation suggesting how much to send or to send back before making any decision. Our second research question addresses how trust and trustworthiness interact with voluntary adherence to this non-binding rule.[5] This research question speaks to the open literature investigating the role of suggested actions,[6] showing that suggested actions do not always bring socially optimal decisions.[7] By introducing an exogenous signal, we measure how ex-ante and/or ex-post information raises adherence to a fair recommendation, avoiding the negative consequences of inequality (Stiglitz, 2012; Verhaeghe, 2014; Pickett and Wilkinson, 2009; Eckel et al., 2022). To follow up with our previous examples, empirical and experimental studies have shown that trust is positively and significantly related to voluntary adherence to certain recommendations if these follow fairness and equality criteria.[8]

Our results suggest that *ex-ante* and *ex-post* information impact senders' decisions in different ways. While senders strongly react to the receiver's *reputation*, trust increases only with *transparency*. This result is

---

[2] See Buchan et al. (2012) for a review of this literature, Fatas et al. (2010), (2015) and (2020), Eckel et al. (2010) for an analysis of behavioural differences triggered by information flows.

[3] See Bolton et al. (2005), Barrera and Buskens (2009), Buskens et al. (2010), Cassar and Rigdon (2011), and Frey et al. (2019).

[4] See Bolton et al. (2005), Barrera and Buskens (2009), Buskens et al., (2010), Cassar and Rigdon (2011), and Frey et al. (2019).

[5] See Fatas et al. (2018) for an analysis of preferences to conform to different signals.

[6] See Cialdini et al. (2006), Espinosa et al (2019), Reidl et al. (2019), Butz and Harbring (2020), Bicchieri et al. (2021), and Restrepo-Plaza and Fatas (2022) for the role of the social norms and suggested actions.

[7] See Dale and Morgan (2010), Charness and Cheung (2013), Edwards and List (2014), Chuan and Samek (2014), Goswami and Urminsky (2016), Reiley and Samek (2019).

[8] See Kasper et al. (2015), Gobena and Van Dijke (2017), Van Dijke et al. (2019).

driven by a large drop in the proportion of participants refusing to send any amount to the receiver in *Transparency*. Overall, our results suggest that e*x-post* social information (as in *Transparency*) is more effective than *ex-ante* aggregate information (as in *Reputation*) in increasing trust. In sharp contrast with previous work, neither *ex-ante* nor *ex*-post information affects receivers' aggregate trustworthiness. In. both conditions, senders and receivers only partially follow the non-binding recommendation, and adherence declines with the amount suggested by the signal. The rest of our paper is organized as follows. Section 2 describes our experimental design and procedures. Section 3 reports the experimental results. Finally, Section 4 concludes this paper.

2. Methods and experimental design

2.1. The game

The standard trust game first introduced by Berg et al. (1995) focuses on interactions between two players. In the first stage, the first mover decides how much of her endowment to send to the second mover. The second mover receives that amount multiplied by a factor k, with $k>1$. In the second stage, the second mover decides how much to send back to the first mover. We implement a variant of the trust game using a three-node networked investment game already studied in this literature (e.g., Buskens et al., 2010; Cassar & Rigdon, 2011; Frey at el., 2019). In the game, two senders interact with the same receiver (see Fig. 1) and simultaneously decide how much of their endowment to invest (*trust*). The second mover receives separately both amounts multiplied by a common factor (four, in our experiment) and decides how much to send back to each sender (*trustworthiness*). Both trust games are separable in the sense that each receiver decision is only constrained by the amount received from each sender.

At the beginning of the experiment, participants go through a trial time, identical to the real-effort task senders will face each round to get their endowment. After the trial phase is finished, the experiment starts. Subjects are randomly assigned to a role (sender or receiver) and a cohort of six participants (four senders and two receivers). For the whole experiment, participants keep their fixed roles and cohort composition does not change. At the beginning of each round, two networks of three players (two senders and one receiver) are randomly formed in each cohort. Senders and receivers within each cohort are randomly rematched in new networks at the beginning of each round keeping their roles (sender or receiver), so senders have a 50% chance of playing with the same receiver in the next round. The game is repeated twenty rounds and complete information about the rules of the game is provided to participants. Table 1 displays the game sequence we describe in detail below.

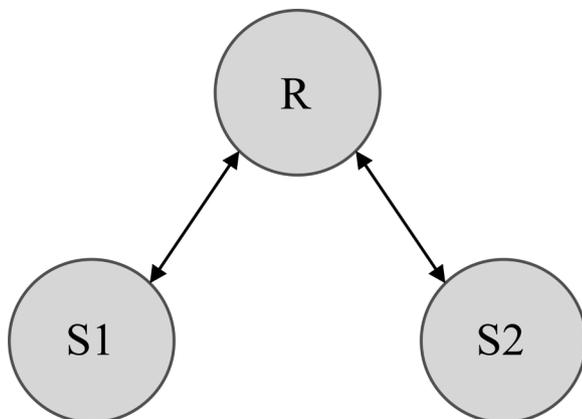*1st stage – Real effort stage*: At the beginning of each round, senders

Table 1
Proceeding of an experimental round.

1st stage: Senders	2nd stage: Senders	3rd stage: Senders	4th stage: Receiver	5th stage: Receiver
Real-Effort Task	Signal 1	Trust (X)	Signal 2	Trustworthiness (Y)

perform a real-effort task (RET) to earn their own endowment.[9] Senders add two-digit numbers for one minute and for each addition correctly solved, their endowment increases by 2 Experimental Currency Units (ECU, up to a maximum of 10 ECU). Receivers do not participate in this task and receive a fixed endowment of 5 ECU.

2nd stage – Signal 1: Each Sender receives a random non-binding signal (see Section 2.2) with a suggested proportion of endowment to send to the receiver (0%, 20%, 40%, 60%, 80%, or 100%).

3rd stage – Sending decision: Each sender j decides a proportion X_j of her endowment to send to the receiver: 0%, 20%, 40%, 60%, 80%, or 100%. The amount is multiplied by four and received by the second mover.

4th stage – Signal 2: The receiver observes the amounts received from both senders and receives two non-binding recommendations, one for each sender (see Section 2.2) about the proportions to return (0%, 20%, 40%, 60%, 80%, or 100%).

5th stage – Returning decision: The receiver decides a proportion Y_j to return to each sender j (0%, 20%, 40%, 60%, 80%, or 100%).

At the end of each period, participants are informed of outcomes. Receivers get information about both trust games (amounts sent and returned), whereas senders only receive information about the actions and outcomes in the trust game they played. The game described above corresponds to our *Baseline treatment*. In two other treatments, additional information is provided to senders. We further describe both information treatments in Section 2.3.

2.2. Design: signals

In our experimental framework, both senders and receivers receive a non-binding recommendation before making their decisions. The signal is presented to the players as a "suggested rule" (D), which is randomly determined each round by a virtual die (see Table 2):

$$S = 0.2 \cdot (D - 1)$$

The signal therefore follows a discrete uniform distribution: S ~ DU (6, 0, 0.2) and every participant is aware of the distribution S from the beginning of the experiment.

If both senders and receivers fully follow the signal throughout the whole experiment, the expected payoff of receivers and senders are identical. Out of their endowment of 10 ECU, senders keep on average 5

Table 2
Signals (outcome of the die) and corresponding recommendation.

Outcome of the die:	1	2	3	4	5	6
% to keep	100%	80%	60%	40%	20%	0%
% to send to the receiver or % to send back to the sender	0%	20%	40%	60%	80%	100%

Fig. 1. The three-node network of two senders (S1 and S2) and one receiver (R).

---

[9] Throughout the experiment, 97% of senders start the round with the maximum endowment of 10 ECU. This design feature has not been implemented to introduce endowment heterogeneity, but rather to give a feeling of entitlement to participants. If endowment was not earned and deserved, participants might have been tempted to consider the trust game as a redistribution game rather than an investment game. See Houser and Xiao (2014) for a discussion of "house money effects" in trust games.

ECU. This amount is matched by the 5 ECU endowment provided to receivers. The 5 ECU investment made by senders becomes 20 ECU in the hands of receivers and is expected to be equally distributed if receivers follow the signal. The non-binding signal could therefore be considered as a "fair rule", in that it equalizes expected payoffs between complying participants. Furthermore, the outcome generates efficiency gains, outperforming the Nash equilibrium outcome in which senders do not send and receivers have nothing to return. The signal outcomes for both senders and receivers were explained to participants with examples (see instructions in appendix).

2.3. Design: information

In the *Baseline treatment*, senders do not receive information about the receiver's past performance before making any decision, and they only receive feedback about the trust game they play at the end of each period. This experimental framework is close to the "partial information" condition studied by Cassar and Rigdon (2011). In the other two treatments, we provide senders with additional information about receivers' actions.

In the *Reputation treatment*, senders receive additional *ex-ante* information at the beginning of each round. This information corresponds to the average proportion returned to senders in the previous round by the receiver they are matched with. Imagine a receiver (Cristina) returning 20% to one sender (Mary) and 40% to the other sender (Joe) in round 1. At the beginning of round 2, Cristina (the receiver) is re-matched with two other senders within her cohort: Paul and Nancy. Paul and Nancy receive information about Cristina's average decision in round 1 (30%) before sending any amount to her in round 2. If Cristina receives 0% from both Mary and Joe in round 1, Paul and Nancy receive information at the beginning of round 2 that Cristina returned 0% on average in round 1. Senders never receive information about the individual amounts sent by Joe and Mary.

In the *Transparency* treatment, senders receive additional *ex-post* information at the end of each round. This information corresponds to the outcome of the other trust game in their network. Imagine a receiver Bob returning 20% to Natalia and 40% to James in round 1. At the end of round 1, Natalia and James receive information about the proportion Bob returned to each of them. If Natalia sent 0% to Bob, James receives information at the end of the round that Bob returned 0% to Natalia. In any case, they do not receive information about the individual amounts sent by James and Natalia.

At the beginning of the experiment, all participants in *Reputation* and *Transparency* are informed of the information disclosed to them in the experiment.

2.4. Procedures

Experimental sessions were run at ESSEXLab (University of Essex). We electronically recruited 108 participants, undergraduate students from different disciplines, all inexperienced in trust games. For each treatment, we collected data from six independent observations (six-participant cohorts). On average, a session lasted 100 min, including initial instructions, quiz, trial phase, final questionnaire, and payment of the subjects. The average payment was around £12.50, including a show up fee of £5. The instructions were read aloud, and the computerized experiment used Z-TREE (Fischbacher, 2007).

2.5. Design: theory and hypotheses

Despite its original features, our experimental design provides subjects with monetary incentives similar to other trust games. Since our game is finitely repeated, the sub-game perfect Nash equilibrium can be solved by backward induction. In the last round, a rational receiver maximizes her profit by keeping the full amount received. Anticipating this, a rational and profit-maximizing sender does not send any amount.

The equilibrium is therefore characterized by no efficiency gains, so payoffs equal initial endowments. The inclusion of partial information in *Reputation* and *Transparency* does not modify this prediction and Nash equilibria are the same in all three treatments. Our theoretical predictions are also unaffected by the presence of a non-binding recommendation, as profit-maximizing agents have no incentive to consider the outcome of the die.

We are aware that a vast existing literature shows that such predictions are not verified empirically. Simple models may indeed overlook decisive behavioral issues. Relying on previous experimental findings, we formulate three behavioral hypotheses to address the behaviorally relevant features of our experimental design. In *Reputation*, information about the receiver's past performance may signal her future decisions, being easier for senders to trust a receiver who returned generously in the previous round, as documented in previous experimental studies (Barrera and Buskens, 2007; Cassar & Rigdon, 2011), creating incentives for receivers to build a positive reputation (Bohnet & Huck, 2004). Altogether, those elements suggest that trustworthiness and trust should increase in *Reputation* relative to *Baseline* (across treatments) and that receivers failing to build a positive reputation will be penalized by senders (*within* treatment).

Hypothesis 1: *Trust and trustworthiness will be higher in Reputation than in Baseline (between treatments effect). Senders will trust receivers with a good reputation more than receivers with a low reputation (within treatment effect).*

In *Transparency*, horizontal comparisons may open the door for trust (mistrust) being rewarded (sanctioned): higher (lower) amounts sent are rewarded (punished) with higher (lower) returns, and senders may be aware of it. A positive correlation between amounts sent and returned is a standard finding in the literature, starting from Berg et al. (1995), in line with the concept of reciprocity. Reciprocity may trigger behavioral responses based on envy and competition between senders (Cassar & Rigdon, 2011; Huck et al., 2012), and senders receiving less might send more in the following round. If senders receiving more do not decrease the amount sent as much as senders receiving less increase theirs, the net effect on trust should be positive in *Transparency* compared to *Baseline* (see Fallucchi et al. (2021)'s model of asymmetric inequality aversion). Reciprocal receivers should send back more to generous senders and, overall, trustworthiness should increase with trust (being the distance between both returning decisions a proxy of receivers' reciprocity).

Hypothesis 2: *Trust and trustworthiness will be higher in Transparency relative to Baseline. Larger inequalities between returning decisions will boost trust.*

We are also interested in understanding how trust and trustworthiness interact with voluntary adherence to the exogenous signals received by participants, who could follow it if perceived as a fair recommendation. As participants may also balance fairness with its cost, they may imperfectly follow it (e.g., sending or returning less if the signal recommends spending most of their resources). Experimental studies of suggested donations (Dale & Morgan, 2010; Charness & Cheung, 2013; Edwards & List, 2014) show that receiving suggestions from the charity is more effective in fostering cooperative behavior than having no recommendations at all, and that donations collapse as the recommendations increase. As the signal in our experiment corresponds to the outcome of a die, rather than a recipient, and its exogenous nature is obvious, our participants may disregard the signal. We predict however that even in this minimal setting, individuals will be biased toward the signal.

Hypothesis 3: *Trust and trustworthiness will be positively but imperfectly correlated to the recommendation derived from the non-binding signal. Low signals (e.g., send/return 0% or 20%) will be followed more frequently than high signals (e.g., send/return 80% or 100%).*

3. Results

Section 3.1 provides an overview of our experimental data, together

with non-parametric tests as a first approach to document treatment effects. Section 3.2 focuses on our information treatments (Hypotheses 1 and 2) and Section 3.3 on how the non-binding signal shapes agents' decisions (Hypothesis 3). Throughout our analysis, we define *trust* as the proportion of endowment invested by the first mover and *trustworthiness* as the proportion of the amount received that is returned by the second mover.[10]

3.1. Results: overview

Table 3 reports summary statistics about decisions of both senders and receivers. Wilcoxon Mann-Whitney tests are run on independent observations (cohorts) to test for significant differences between treatments.

The upper panel of Table 3 displays decisions from 72 senders both in aggregate and for each treatment. The first column presents the average level of trust. Trust is on average higher in the *Transparency treatment* than in the *Baseline treatment* ($p = 0.0372$) and the *Reputation treatment* ($p = 0.0483$), and there is no significant difference between *Baseline* and *Reputation* ($p = 0.2461$).

The second column only considers instances in which trust was positive, excluding interactions in which the first mover decided to invest nothing. Results indicate that sender decisions do not differ significantly between treatments in this case, suggesting that the higher

Table 3
Descriptive statistics.

	Trust (Proportion X sent by senders)		
	Trust (Including $X = 0$)	Trust ($X > 0$ only)	Proportion of $X = 0$
Total	0.3879	0.5093	0.2384
	(0.3213)	(0.2715)	(0.1467)
Baseline	0.3795	0.5050	0.2484
	(0.3206)	(0.2707)	(0.1075)
Reputation	0.3454	0.4848	0.2875
	(0.3165)	(0.2700)	(0.2092)
Transparency	0.4388 **	0.5345	0.1792 *
	(0.3206)	(0.2720)	(0.0571)

	Trustworthiness (Proportion Y returned by receivers)		
	Trustworthiness (Including $X = 0$)	Trustworthiness ($X > 0$ only)	Proportion of $Y = 0$ ($X > 0$ only)
Total	0.2518	0.3305	0.2498
	(0.2839)	(0.2825)	(0.0771)
Baseline	0.2833	0.3767	0.2410
	(0.3156)	(0.3118)	(0.0610)
Reputation	0.2288	0.3211	0.2544
	(0.2697)	(0.2692)	(0.0929)
Transparency	0.2433	0.2964	0.2538
	(0.2613)	(0.2596)	(0.0706)

Notes: Standard deviations are displayed in parentheses. Stars report significance level from Wilcoxon Mann-Whitney tests run on independent observations (cohorts of 6 participants) to confirm differences with the Baseline treatment;.
* $p < 0.10$,.
** $p < 0.05$, *** $p < 0.01$. We excluded from the analysis of the receivers (bottom part) all the interactions in which receivers received 0% from a sender and had no other choice than returning 0%.

level of trust observed in *Transparency* is mainly a result of a lower proportion of senders displaying no trust at all. The third column of Table 3 partially supports this assertion as the proportion of interactions in which no trust was displayed is lower in *Transparency* (17.92%) than in *Baseline* (24.84%) and *Reputation* (28.75%). This difference is however not statistically significant at the 5% level.

Additionally, substantial differences between treatments were observed in the frequency of instances where receivers were presented with no choice (i.e., both senders sent nothing). This occurred in 7 out of 240 observations (2.92%) in *Transparency*, 15 out of 240 observations (6.25%) in *Baseline* and was much higher in *Reputation* with 32 out of 240 observations (13.33%).[11] It is worth noting that receivers collecting nothing from senders had an extremely low reputation score in *Reputation:* 0.023 (they returned on average 2.3% of the amount received in the previous period).

The lower panel of Table 3 reports returning decisions by 36 receivers. Considering only cases in which receivers received a positive amount from senders ($X > 0$), trustworthiness in *Transparency* and *Reputation* is lower than in *Baseline*, being differences are not statistically significant ($p = 0.3367$ and $p = 0.1495$ respectively). Overall, and consistent with reciprocity (McCabe et al., 2003), receivers returned on average 25.08% of the amount received, which is equivalent to the proportion invested by senders (as the amount sent was multiplied by four).

3.2. Results: information

We implement random effects ordered Probit models to study both trust and trustworthiness. This model is well-suited to the discrete nature of our dependent variables, and allows us to capture potential non-linearity between decisions.[12] The model considers an unobserved continuous latent variable $y_{it}^* = \beta' x_{it} + u_i + \varepsilon_{it}$ with x_i a vector of explanatory variables, β the corresponding parameter vector, u_i a random individual effect and ε_{it} an error term. The unobserved latent variable can be related to the observed ordered variable y_{it} (either trust or trustworthiness) through a threshold mechanism which divides the latent variable in J intervals by using $J + 1$ threshold parameters τ_0, τ_1, ..., τ_J. The probability that our interest variable takes value j can then be expressed as (being Φ the CDF of the normal distribution):

$$P(y_{it} = j | x_{it}) = \Phi(\tau_j - \beta' x_{it}) - \Phi(\tau_{j-1} - \beta' x_{it})$$

Tables 4 and 5 present the results of four models each. Column (1) reports estimates from a pooled regression, considering all our experimental data, introducing treatment dummies. Column (2) focuses on participants in *Baseline*, column (3) focuses on participants in *Reputation* and column (4) focuses on *Transparency*. We include in all regressions the outcome of the random non-binding signal, as thoroughly explored in Section 3.3, a period variable and basic demographic controls (gender, age, British nationality).

In Table 4, focused on *Trust*, model (3) includes the *reputation score of receivers* observed by senders before their decision, and a *receiver with no*

---

[10] These labels denominations are common in the trust game literature, but it is nevertheless important to define this concept correctly. An increasing "trust" in our findings would not necessarily imply that the intrinsic trust of the participants increases. The investment decision could be driven by other factors, such as the presence of information that relieves ambiguity, or the willingness to compete with another sender.

[11] Information regarding reputation affects the amounts straightforwardly received by the receivers. Both senders receive the same information regarding the decisions of the receiver in the previous round. If the reputation of the receiver is very low, it may be interpreted as a signal that both senders act in accordance, by sending nothing. This mechanism cannot occur in *Baseline*, where no additional information is provided, or in *Transparency*, where the information is displayed after all decisions took place. Our results are in line with recent literature suggesting that past outcomes and reputation affect the perception of others and expectation of future behaviour, which at the same time shape preferences (see Dezső, et al., 2015; Dezső, et al., 2019; Dezső, et al., 2022).

[12] Increasing trust from 0% to 20% may indeed not be equivalent to increasing trust from 20% to 40%.

Table 4
Random effects ordered Probit regressions for trust.

	Dependent variable: Trust			
	(1) Pooled	(2) Baseline	(3) Reputation	(4) Transparency
Baseline treatment	Ref.	–	–	–
Reputation treatment	−0.063	–	–	–
	(0.258)			
Transparency treatment	0.304**	–	–	–
	(0.125)			
Reputation of the receiver	–	–	1.196**	–
			(0.542)	
Receiver with no reputation [a]	–	–	−0.663	–
			(0.538)	
Received less than the other sender	–	–	–	0.299**
				(0.138)
Received more than the other sender	–	–	–	0.090
				(0.142)
Non-binding signal	0.138***	0.102*	0.167***	0.164**
	(0.032)	(0.056)	(0.046)	(0.066)
Trust in the previous period	0.811***	1.280***	0.239	0.870**
	(0206)	(0.234)	(0.363)	(0.395)
Trustworthiness experienced in the previous period	0.638***	0.652***	0.399*	0.736***
	(0.088)	(0.083)	(0.205)	(0.170)
Period	−0.016**	−0.032***	−0.012	0.000
	(0.007)	(0.009)	(0.012)	(0.003)
Female	−0.237	−0.163	−0.477	−0.290
	(0.178)	(0.256)	(0.350)	(0.240)
Age	0.041	0.013	0.055	0.021
	(0.033)	(0.036)	(0.057)	(0.034)
British	−0.066	−0.196	−0.180	0.193
	(0.204)	(0.315)	(0.311)	(0.434)
# Observations	1366	454	456	456
# Groups	72	24	24	24

Random effects are at the individual level. Standard errors (in parentheses) are clustered on the independent observation level (cohorts of six individuals);.

* $p < 0.10$,.
** $p < 0.05$,.
*** $p < 0.01$.

[a] Receivers that did not receive any endowment in previous period (i.e., received 0% from both senders) could not make any choice and therefore did not build any reputation.

reputation dummy to capture instances in which receivers had no chance to display trustworthiness in the previous period (as they received nothing from senders). Model (4) includes two dummy variables characterizing the amounts received by senders (*received less than the other sender* and *received more than the other sender*). Dynamic effects are captured by *trust in the previous period* and by *trustworthiness experienced in the previous period*. In Table 5, focused on *Trustworthiness*, we include trust exhibited by both senders (Cassar & Rigdon, 2011),[13] and dynamic effects are captured by *total trust in the previous period* (the sum of trust decisions observed in the previous period) and by *total trustworthiness experienced in the previous period* (the sum of both returning decisions in the previous period). We discuss the results, treatment by treatment, after both tables.

3.2.1. Results: reputation

In *Reputation*, senders observe the average lagged trustworthiness of their receiver (their reputation score). Hypothesis 1 states that trust and trustworthiness should be higher in this treatment than in *Baseline*, as higher reputation would lead to higher trust. In sharp contrast with hypothesis 1, the previous subsection showed that trust does not significantly increase. Column (1) of Table 4 confirms this finding, as the coefficient associated to the reputation treatment dummy is not

---

[13] In networked trust games with two senders and one receiver, Cassar and Rigdon find evidence of "comparative trust": the return decisions by receivers depend in part on the investment behaviour along the other link in the network.

Table 5
Random effects ordered Probit regressions for trustworthiness.

	Dependent variable: Trustworthiness			
	(1) Pooled	(2) Baseline	(3) Reputation	(4) Transparency
Baseline treatment	*Ref.*	–	–	–
Reputation treatment	−0.205	–	–	–
	(0.326)			
Transparency treatment	−0.256	–	–	–
	(0.238)			
Trust from the sender	0.411*	−0.0367	0.755***	0.671*
	(0.219)	(0.393)	(0.170)	(0.376)
Trust from the other sender	0.016	0.137	−0.477**	0.176
	(0.122)	(0.193)	(0.230)	(0.147)
Non-binding signal	0.220***	0.164***	0.225***	0.283***
	(0.029)	(0.043)	(0.043)	(0.064)
Total trust in the previous period	0.044	−0.124	0.303***	−0.019
	(0.100)	(0.192)	(0.101)	(0.196)
Total trustworthiness in the previous period	0.156**	0.160*	0.213	0.176
	(0.066)	(0.084)	(0.170)	(0.147)
Period	−0.016**	−0.006	−0.031***	−0.019
	(0.007)	(0.005)	(0.006)	(0.021)
Female	0.390	0.645	0.937	−0.288
	(0.324)	(0.689)	(0.735)	(0.064)
Age	0.025	−0,019	0.159	0.101
	(0.022)	(0.032)	(0.273)	(0.108)
British	−0.272	0.427	−0.701	0.067
	(0.317)	(0.513)	(0.750)	(0.504)
# Observations	1041	342	323	376
# Groups	36	12	12	12

Random effects are at the individual level. Standard errors (in parentheses) are clustered on the independent observation level (cohorts of six individuals);.

* $p < 0.10$,.
** $p < 0.05$,.
*** $p < 0.01$. **Note**: we excluded from the analysis in all the Columns the interactions in which receivers received 0% from a sender and had no other choice than returning 0%.

significantly different than zero. However, this result does not imply that senders ignore the information embedded in the reputation score. Column (3) of Table 4, only with participants in *Reputation,* show that the reputation of the receiver has a positive and significant effect on trust.

To give a quantitative estimate of this effect, we build from the coefficients reported in column (3) of Table 4 and calculate the predicted probabilities for a sender to either show no trust at all or to show a complete trust. Those probabilities are calculated as:

$$\begin{cases} P(trust = 0 | X) = \Phi(\tau_1 - X\widehat{\beta}) \\ P(trust = 1 | X) = \Phi(1 - \tau_5 - X\widehat{\beta}) \end{cases}$$

With X a vector of characteristics for which all covariates are set to their average value, except for the reputation which we vary to identify its effect on trust. Φ is the CDF of the normal distribution, $\widehat{\beta}$, τ_1 and τ_5 are parameters estimated from the random-effects probit reported in Column (3) of Table 4. Table 6 reports the estimated probabilities, as well as the outcome of statistical tests comparing the predicted probability with a zero probability. Our results indicate that, relative to a reputation score of 0%, a maximum reputation score of 100% would reduce by more than 80% the likelihood of the sender displaying no trust at all (from 31.78% to 5.74%). Similarly, the probability of full trust (sending the receiver the full amount) increases dramatically with the receiver's reputation (from 3.56% to 23.57%). Even when no significant differences are observed across treatments (*Reputation* versus *Baseline*), reputation scores make a significant difference in the *Reputation* condition, as senders modulate their trust (the amount they send) following the information received about the reputation of their receiver.

One could argue that the impact of reputation could reflect the interaction dynamics within the cohort and not the effect of reputation disclosure at the beginning of the period. To explore this possibility, we compare the impact of reputation (as disclosed to senders) in *Reputation*,

Table 6
Distribution of receivers' reputation and predicted trust.

Reputation	Frequency	Predicted probability for no trust ($X = 0$)	Predicted probability for full trust ($X = 1$)
0%	33	0.3178***	0.0356*
10%	30	0.2798***	0.0450**
20%	36	0.2440***	0.0562**
30%	27	0.2109***	0.0696***
40%	28	0.1805***	0.0852***
50%	28	0.1530***	0.1034***
60%	9	0.1284***	0.1243***
70%	3	0.1067**	0.1479***
80%	2	0.0877*	0.1744***
90%	2	0.0713	0.2037***
100%	0	0.0574	0.2357***

Note: The reputation of the receiver is displayed to two senders but reported only once in this table. Predicted probabilities are computed for average values of every variable except *reputation of the receiver* and *no reputation*.

* $p < 0.10$,.
** $p < 0.05$,.
*** $p < 0.01$.

with the impact of reputation (undisclosed to senders) in *Baseline* and *Transparency*. While trust is positively correlated to reputation scores in *Reputation* ($\rho = 0.245$, $p < 0.001$), it is not in *Baseline* ($\rho = 0.068$, $p = 0.157$) nor in *Transparency* ($\rho = 0.053$, $p = 0.269$).

The lack of significant differences between *Reputation* and *Baseline* in the amounts sent (trust) is consistent with receivers failing to recognize the positive correlation between reputation scores and trust. In the previous subsection we documented how trustworthiness does not increase on average in *Reputation* compared to *Baseline*. Column (1) of Table 5 confirms this finding, since the coefficient associated to the reputation treatment dummy is not significantly different from zero. Column (3) of Table 5 focus on participants in *Reputation* and, consistent with the comparative trust hypothesis formulated by Cassar and Rigdon (2011), shows how trustworthiness toward each sender follows both senders' trust. While trusting more makes the receiver more generous with senders (as suggested by the large and positive coefficient of *Trust from the sender*: +0.755), low trust by the other sender makes your own level of trust more generous (as the negative and significant coefficient for *Trust from the other sender* suggests: −0.477). In other words, returning decisions by receivers are based on the relative, rather than absolute, trust exhibited by both senders.

Although trustworthiness comes at a cost, building up reputation pays off for receivers in the long run (i.e., over the course of the 20 periods): overall earnings are positively correlated to their average reputation score ($\rho = 0.286$, $p < 0.001$) only in *Reputation*, and not in the other treatments. As reputation is not disclosed in the other conditions, the average reputation score is significantly and negatively correlated to earnings in *Baseline* ($\rho = -0.7712$, $p < 0.001$) and *Transparency* ($\rho = -0.6245$, $p < 0.001$ in the treatment).

We summarize our findings so far in our first result:

Result 1: *Trust increases with receivers' reputation if disclosed (as in the Reputation condition). Receivers who build-up their reputation are better off than receivers with low reputation. On average, however, trustworthiness does not increase relative to the Baseline, and trust is not significantly higher on average in the Baseline treatment.*[14]

3.2.2. Results: transparency

In *Transparency*, senders observe the trustworthiness of a receiver interacting with both senders. Hypothesis 2 states that trust and trustworthiness will be higher than in *Baseline*, particularly if receivers discipline senders returning unequal amounts (and giving them incentives to trust more). In Section 3.1 we documented how trust increases on average when transparency is introduced (as an average treatment effect, see Table 3), in line with hypothesis 2. Column (1) of Table 4 confirms this finding, as the coefficient of the *Transparency treatment* dummy is significantly positive.

Column (4) of Table 4 focuses on the sample of participants taking part in *Transparency*. Senders rewarded with a high level of trustworthiness in the previous period increase trust (as the positive and significant coefficient of *Trustworthiness experienced in the previous round* suggests: +0.736), and senders in an advantageous position (trusting more in the previous round), do not trust less but more (as the positive and significant coefficient of *Trust in the previous round* shows: +0.870). In *Transparency*, senders receiving less than the other participant increase their trust by 6 percentage points on average.[15]

In contrast, trustworthiness does not significantly increase, on average, in *Transparency*, relative to *Baseline*, as discussed in Section 3.1, and confirmed in our econometric estimates (see Column (1) of Table 5). This finding is not in line with hypothesis 2, and not necessarily surprising. First and foremost, a receiver does not directly benefit from higher trustworthiness, as they will not interact with the same senders in the next period. Second, the social comparison mechanism triggered in *Transparency* does not call for higher average trustworthiness, but for different returning decisions. Taken together both ideas, the incentives to exhibit trustworthiness in *Transparency* are modest. Overall, our experimental data partially support hypothesis 2. We summarize our findings in our Result 2:

Result 2: *In Transparency, trust is higher on average than in Baseline and Reputation, but trustworthiness is not.*

3.3. Results: following the signal

Hypothesis 3 states that senders and receivers should imperfectly follow the signal they observe. All regression estimates reported in Tables 4 and 5 provide strong support for this hypothesis, regardless of the treatment considered, as the coefficient associated to the non-binding signal is systematically positive and significant, regardless of whether we estimate trust or trustworthiness. We take a closer look at the data to fully explore and test hypothesis 3, identifying whether how participants follow different non-binding signals.

Senders fully follows the signal 30% of the times in *Baseline*, 33% in *Reputation* and 30% in *Transparency*. Similarly, receivers mimic the signal they observe 34% in *Baseline*, 34% in *Reputation* and 36% of the times in *Transparency*. We do not observe any significant treatment difference in the likelihood for a participant to fully follow the signal.

Fig. 2 reports the distribution of trust for each signal value. The main message from Fig. 2 is that participants conform to the signal if not too demanding: signals suggesting sending between 0% and 60% of their endowment make participants to conform frequently (the value of the signal is the modal outcome in all four cases). However, when the outcome of the die is higher than 60%, the modal trust decision is 0%. This observation echoes previous experimental findings (Dale & Morgan, 2010) in the context of public good games showing that non-binding recommendations shape decisions if perceived as "reasonable". Along the same line, field experiments suggest that higher suggestions induce less generous contributions (Charness & Cheung, 2013; Edwards & List, 2014), and that conformity is context dependent (Fatas

---

[14] Note that our experimental setting might underestimate the impact of reputation, as subjects interact with different participants in cohorts of size six. The relatively small matching group size makes possible for receivers in *Baseline* to build up trust over time. A way to circumvent this limit would have been to implement a perfect-stranger matching protocol reducing the number of independent observations.

[15] This figure is the average increase in trust observed among senders who were in a disadvantageous position in the previous period. It does not rely on our econometric estimation.
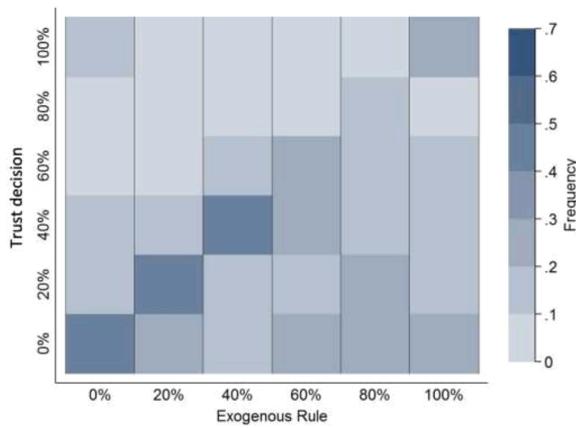
Fig. 2. . Frequencies of trust decisions regarding exogenous signal.

et al., 2018).

Our findings are not different when focusing on trustworthiness.
Fig. 3 reports the distribution of receivers' decisions for each signal
value.[16] Whereas suggestions to return relatively small amounts are
often followed, only a few receivers comply with more demanding sig-
nals, suggesting more than 50%. When the outcome of the die suggests
returning everything (100%), receivers do so in only 14.87% of cases,
certainly above the prevalence of the same decision (return everything)
when the suggestion is different (2.55% of cases), but also below the
conformity observed for more reasonable and less demanding
suggestions.

Table 7 reports estimates from random-effect Probit regressions
aiming at identifying what brings individuals to follow the signal. To test
for between-treatment differences, we introduce treatment dummies in
these models, and we include the signal value as a covariate to confirm
that the likelihood to follow the suggestion decreases with its value. All
models also include *period* variables and the same demographic infor-
mation (*Female, Age* and *British*) used in other regressions as control
variables.

Column (1) of Table 7 uses trust as dependent variable. Consistent
with our previous findings, the likelihood to follow the signal signifi-
cantly decreases with its value (coefficient: −0.160). We also observe
that the likelihood of following the signal slightly and significantly de-
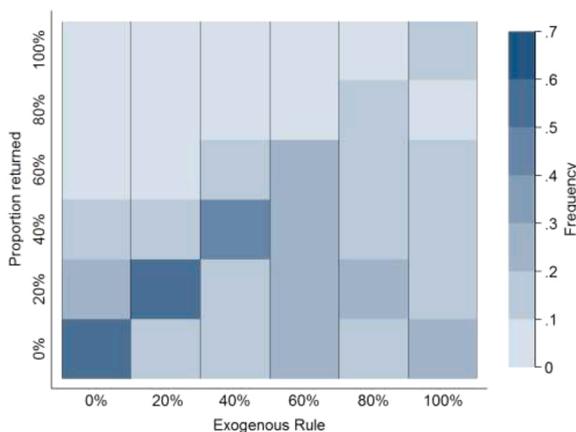creases over time (coefficient: −0.015). Our estimates confirm the lack

Fig. 3. . Frequencies of returning decisions regarding exogenous signal.

Table 7

Random-effect probit regression for following the suggested signal.

	Dependent variable: matches the signal (dummy)	
	(1) Senders	(2) Receivers
Baseline Treatment	*Ref.*	*Ref.*
Reputation Treatment	0.066	−0.0553
	(0.102)	(0.179)
Transparency Treatment	−0.017	0.012
	(0.133)	(0.177)
Non-binding signal	−0.160***	−0.335***
	(0.0326)	(0.0255)
Period	−0.015**	−0.0157**
	(0.007)	(0.006)
Female	0.091	0.270
	(0.124)	(0.134)
Age	−0.053**	−0.032
	(0.025)	(0.014)
British	−0.064	−0.134
	(0.121)	(0.103)
Intercept	1.227**	1.267***
	(0.523)	(0.369)
# Observations	1440	1096
# Individuals	72	36

Random-effects are at the individual level. Standard errors (in parentheses) are
clustered on the independent observation level (cohorts of six individuals); * $p <$
0.10,.
** $p < 0.05$,.
*** $p < 0.01$. **Note**: we excluded from the analysis of the receivers (Column 2)
all the interactions in which receivers received 0% from a sender and had no
other choice than to return 0%.

of treatment effects, as suggested by the descriptive analysis discussed
above. As Column (2) of Table 7, using trustworthiness as a dependent
variable, suggests, the likelihood to follow the signal also decreases with
its value for receivers (coefficient: −0.335), and adherence to the signal
also decreases with time (coefficient: −0.0157). As in Column (1), no
significant treatment differences are observed. We summarize our
findings in our next result:

Result 3: *For both senders and receivers and across all treatments, sig-
nals have a significant and positive impact on individual decisions, but the
likelihood of following signals decrease with its value and with time.*

We can only speculate on precise explanations for our last finding. As
following the signal may be considered as costly and risky, the higher
the signal value, the higher the perceived cost. One may also argue that
low signals recommendation may be interpreted as a moral wiggle room,
as (assuming that trust is considered by participants as the "moral thing
to do") a low recommendation may lower the moral cost for participants
to send a low amount.[17] Our experimental design unfortunately does not
allow us to disentangle between those channels.

4. Conclusion

We run a laboratory experiment to investigate trust and trustwor-
thiness in a networked investment game in which two senders interact
with a receiver. Previous studies using a comparable framework
(Buskens et al., 2010; Cassar & Rigdon, 2011; Frey et al., 2019) docu-
mented the positive effect of full information on actions and outcomes
(increasing both trust and trustworthiness). In this paper, we study the
more realistic setting in which disclosure of partial information may
have an impact on trust, trustworthiness, and efficiency.

We explore which type of information could be effective in raising all
three variables by comparing a *Baseline treatment* in which senders are
only informed about the actions and outcomes of their own investment
games with two information treatments. In *Reputation*, senders receive

---

[16] We excluded from the analysis all the interactions in which receivers
received 0% from a sender and had no other choice than to return 0%.

[17] We thank an anonymous referee for this suggestion.

ex-ante information regarding the average trustworthiness of the receiver in the previous period. On average, and against Hypothesis 1, neither trust nor trustworthiness increases relative to *Baseline*. However, and in line with the within treatment part of Hypothesis 1, receivers with a high reputation score significantly increase senders' trust. In other words, information allowing for reputation building benefits trustworthy receivers with higher trust (and punishes receivers with a bad reputation). Although we do not observe an increase in trustworthiness compared to *Baseline, ex-ante* information like the one disclosed in *Reputation* may still be a way to increase trust if low reputation receivers may be removed from office (e.g., in local elections; not tested in our experiment).

In *Transparency*, senders receive *ex-post* information about receivers' decisions. In line with Hypothesis 2, trust significantly increases in *Transparency* relative to *Baseline* and *Reputation*. Interestingly, trustworthiness does not vary, boosting receivers' profits. As predicted by Hypothesis 2, a sender who has been disadvantaged (receiving less than the other sender), trusts more in the following round, whereas no change in trust is observed for those who received an equal or larger proportion than their counterpart. Our results are consistent with the idea that information does not need to alter receivers' behavior to benefit society or organizations. More transparency may change the way senders (e.g., citizens or employees) behave because they anticipate that the amounts returned may not be equal, even with no explicit information about the amount sent by the other sender. In other words, receivers' decisions in *Transparency* promote trust even when they are not different on average from the decisions made by receivers in other treatments.

We also study the role played by an exogenous, fair, non-binding and private recommendation, guaranteeing equal expected payoffs. Even when (and in line with our Hypothesis 3), the suggested signal has a significant and positive impact on trust and trustworthiness, it declines with its value and with time in a very similar manner across all treatments. The fact that participants voluntarily comply with the signal echoes past experimental findings suggesting that equality fosters social capital (trust, honesty, cooperation) and reinforces institutions (Fehr et al., 2018; Pickett & Wilkinson, 2010). Our results are also consistent with the idea that a significant determinant of (voluntary) adherence to non-binding signals is the perception of rules as fair (Kasper et al., 2015; Gobena and Dijke; 2017), in line with previous studies on donation suggestions (Karlan & List, 2007). As discussed above, and in line with the final part of our Hypothesis 3, our data suggests that following the signal significantly decreases as the amount suggested by this signal increases, consistent with Dale and Morgan (2010), Charness and Cheung (2013) and Edwards and List (2014).

In summary, our paper provides insights on the role played by information in three-player investment games. Given our relatively small sample size, future research is needed not only to replicate our findings, but to extend our design by adding other relevant features, like a partner matching design (investigating behavioral network dynamics when stable, long-term relationships are possible). Even when it would be too speculative to derive specific lessons for policy design, our study contributes to the growing literature on government transparency reviewed by Michener (2019). Organizations may benefit from non-binding suggestions linked to corporate values, even if they are not part of employees' incomplete contracts. Non-binding signals may help organizations to create more favorable workplace dynamics. As our data strongly documents, reasonable non-binding rules may nudge employees to follow desirable behavior codes, promoting a more reciprocal work environment. Our findings also provide insights on how the link between informational policies and trust in government is mediated by government performance, as in Alessandro et al. (2021): overperforming receivers (governments) may benefit from higher levels of trust by senders (citizens), as in *Reputation*, while social comparisons may trigger trust if information is transparent, as in *Transparency*. In the absence of competitive mechanisms to replace low performing receivers (public officials with a poor reputation record), only transparent information make governments (receivers) to deserve trust.

Data availability

Data will be made available on request.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.socec.2023.102061.

References

Alessandro, M., Lagomarsino, B. C., Scartascini, C., Streb, J., & Torrealday, J (2021). Transparency and trust in government. Evidence from a survey experiment. *World Development, 138*, Article 105223.

Alm, J., Cherry, T., Jones, M., & McKee, M. (2010). Taxpayer information assistance services and tax compliance behavior. *Journal of Economic Psychology, 31*(4), 577–586.

Austin, R.D. (2013). *Measuring and managing performance in organizations*. Addison-Wesley.

Barrera, D., & Buskens, V. (2009). *Third-party effects. eTrust: Forming Relationships in the Online World*, 37–72.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122–142.

Bicchieri, C., Fatas, E., Aldama, A., Casas, A., Deshpande, I., Lauro, M., ..., et al. (2021). In science we (should) trust: Expectations and compliance across nine countries during the COVID-19 pandemic. *PloS One, 16*(6), Article e0252892.

Bohnet, Iris, & Huck, Steffen (2004). Repetition and reputation: Implications for trust and trustworthiness when institutions change. *American Economic Review, 94*(2), 362–366.

Bolton, G. E., Katok, E., & Ockenfels, A. (2005). Cooperation among strangers with partial information about reputation. *Journal of Public Economics, 89*(8), 1457–1468.

Brown, S., Gray, D., McHardy, J., & Taylor, K. (2015). Employee trust and workplace performance. *Journal of economic behavior & organization, 116*, 361–378.

Buchan, N., Fatas, E., & Grimalda, G. (2012). Connectivity and cooperation. *The Oxford Handbook of Economic Conflict Resolution*, 151–181.

Buskens, V., Raub, W., & Van der Veer, J (2010). Trust in triads: An experimental study. *Social Networks, 32*(4), 301–312.

Butz, B., & Harbring, C. (2020). Donations as an incentive for cooperation in public good games. *Journal of Behavioral and Experimental Economics, 85*, Article 101510.

Calabuig, V., Fatas, E., Olcina, G., & Rodriguez-Lara, I. (2016). Carry a big stick, or no stick at all: Punishment and endowment heterogeneity in the trust game. *Journal of Economic Psychology, 57*, 153–171.

Cassar, A., & Rigdon, M. (2011). Trust and trustworthiness in networked exchange. *Games and Economic Behavior, 71*(2), 282–303.

Castro, M. F., & Guccio, C. (2019). Birds of a feather flock together: Trust in government, political selection and electoral punishment. *Public choice*, 1–25.

Charness, G., & Cheung, T. (2013). A restaurant field experiment in charitable contributions. *Economics Letters, 119*(1), 48–49.

Chuan, A., & Samek, A. S. (2014). Feel the Warmth" glow: A field experiment on manipulating the act of giving. *Journal of Economic Behavior & Organization, 108*, 198–211.

Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., & Winter, P. L. (2006). Managing social norms for persuasive impact. *Social influence, 1*(1), 3–15.

Dale, D. J., & Morgan, J. (2010). *Silence is golden: Suggested donations in voluntary contribution games*. Berkeley, USA: University of California.

Dezső, L., Alm, J., & Kirchler, E. (2022). Inequitable wages and tax evasion. *Journal of Behavioral and Experimental Economics, 96*, Article 101811.

Dezső, L., Loewenstein, G., Steinhart, J., Neszveda, G., & Szászi, B. (2015). The pernicious role of asymmetric history in negotiations. *Journal of Economic Behavior & Organization, 116*, 430–438.

Eckel, C. C., Fatas, E., & Wilson, R. (2010). Cooperation and status in organizations. *Journal of Public Economic Theory, 12*(4), 737–762.

Eckel, C. C., Fatas, E., & Kass, M. (2022). Sacrifice: An experiment on the political economy of extreme intergroup punishment. *Journal of Economic Psychology, 90*, Article 102486.

Edwards, J. T., & List, J. A. (2014). Toward an understanding of why suggestions work in charitable fundraising: Theory and evidence from a natural field experiment. *Journal of Public Economics, 114*, 1–13.

Espinosa, M. P., Fatas, E., & Ubeda, P. (2019). Linguistic diversity and out-group discrimination in bilingual societies. *Journal of Behavioral and Experimental Economics, 81*, 102–127.

Fallucchi, F., Fatas, E., Kölle, F., & Weisel, O. (2021). Not all group members are created equal: Heterogeneous abilities in inter-group contests. *Experimental Economics, 24*(2), 669–697.

Fan, X., & Yang, F. (2019). Strategic promotion, reputation, and responsiveness in bureaucratic hierarchies. *Journal of Theoretical Politics, 31*(3), 286–307.

Fatas, E., Heap, S. P. H., & Arjona, D. R. (2018). Preference conformism: An experiment. *European Economic Review, 105*, 71–82.

Fatas, E., Meléndez-Jiménez, M. A., & Solaz, H. (2010). An experimental analysis of team production in networks. *Experimental Economics, 13*(4), 399–411.

Fatas, E., Meléndez-Jiménez, M. A., Morales, A. J., & Solaz, H. (2015). Public goods and decay in networks. *SERIEs, 6*(1), 73–90.

Fatas, E., Meléndez-Jiménez, M. A., & Solaz, H. (2020). Social hierarchies: A laboratory study on punishment patterns across Networks. *Economic Inquiry, 58*(1), 104–119.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171–178.

Frey, V., Buskens, V., & Corten, R. (2019). Investments in and returns on network embeddedness: An experiment with trust games. *Social Networks, 56*, 81–92.

Gobena, L. B., & Van Dijke, M. (2017). Fear and Caring: Procedural Justice, Trust, and Collective Identification as Antecedents of Voluntary Tax Compliance. *Journal of Economic Psychology*.

Goswami, I., & Urminsky, O. (2016). When should the ask be a nudge? The effect of default amounts on charitable donations. *Journal of Marketing Research, 53*(5), 829–846.

Granovetter, M. S. (1985). Economic Action and Social Structure: The Problem of Embeddedness. *American journal of sociology, 91*(3), 481–510.

Houser, D., & Xiao, E. (2014). House money effects on trust and reciprocity. *Public Choice*, 1–13.

Huck, S., Lünser, G. K., & Tyran, J. R. (2012). Competition fosters trust. *Games and Economic Behavior, 76*(1), 195–209.

Karlan, D., & List, J. A. (2007). Does price matter in charitable giving? Evidence from a large-scale natural field experiment. *The American Economic Review, 97*(5), 1774–1793.

Kasper, M., Kogler, C., & Kirchler, E. (2015). Tax policy and the news: An empirical analysis of taxpayers' perceptions of tax-related media coverage and its impact on tax compliance. *Journal of Behavioral and Experimental Economics, 54*, 58–63.

Keefer, P., & Scartascini, C. (2022). Trust: The Key to Social Cohesion and Growth in Latin America and the Caribbean, Inter American Bank of Development.

Lee, D., & Van Ryzin, G. G. (2020). Bureaucratic reputation in the eyes of citizens: An analysis of US federal agencies. *International Review of Administrative Sciences, 86*(1), 183–200.

McCabe, K. A., Rigdon, M. L., & Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization, 52*(2), 267–275.

Michener, G. (2019). Gauging the impact of transparency policies. *Public Administration Review, 79*(1), 136–139.

Ostrom, E., Gardner, R., Walker, J., & Walker, J. (1994). *Rules, games, and common-pool resources*. University of Michigan press.

Pickett, K., & Wilkinson, R. (2010). *The spirit level: Why equality is better for everyone*. Penguin UK.

Reiley, D., & Samek, A. (2019). Round giving: A field experiment on suggested donation amounts in public-television fundraising. *Economic Inquiry, 57*(2), 876–889.

Restrepo-Plaza, L., & Fatas, E. (2022). When ingroup favoritism is not the social norm a lab-in-the-field experiment with victims and non-victims of conflict in Colombia. *Journal of Economic Behavior & Organization, 194*, 363–383.

Stiglitz, J.E. (2012). *The price of inequality: How today's divided society endangers our future*. WW Norton & Company.

Torgler, B. (2007). *Tax compliance and tax morale: A theoretical and empirical analysis*. Edward Elgar Publishing.

Van Dijke, M., Gobena, L., & Verboon, P. (2019). Make me want to pay. A three-way interaction between procedural justice, distributive justice, and power on voluntary tax compliance. *Frontiers in Psychology, 10*, 1632.