Embezzlement and Guilt Aversion

Giuseppe Attanasi, Claire Rimbaud, Marie Claire Villeval

GATE LYON SAINT-ÉTIENNE
UMR 5824

93, chemin des Mouilles
69130 Ecully - France

Maison de l'Université, Bâtiment B
10, rue Tréfilerie
42023 Saint-Etienne cedex 02 - France

http://www.gate.cnrs.fr
gate@gate.cnrs.fr

WP 1807 – April 2018

Embezzlement and Guilt Aversion

Giuseppe Attanasi, Claire Rimbaud, Marie Claire Villeval

Abstract:

Donors usually need intermediaries to transfer their donations to recipients. A risk is that donations can be embezzled before they reach the recipients. Using psychological game theory, we design a novel three-player Embezzlement Mini-Game to study whether intermediaries suffer from guilt aversion and whether guilt aversion toward the recipient is stronger than toward the donor. Testing the predictions of the model in a laboratory experiment, we show that the proportion of guilt-averse intermediaries is the same irrespective of the direction of the guilt. However, structural estimates indicate that the effect of guilt on behaviour is higher when the guilt is directed toward the recipient.

Keywords:
Embezzlement, Dishonesty, Guilt Aversion, Psychological Game Theory, Experiment

JEL codes:
C91

Working Paper

Embezzlement and Guilt Aversion

G. Attanasi[a], C. Rimbaud[b], M.C. Villeval[c]

April 25, 2018

Abstract: Donors usually need intermediaries to transfer their donations to recipients. A risk is that donations can be embezzled before they reach the recipients. Using psychological game theory, we design a novel three-player Embezzlement Mini-Game to study whether intermediaries suffer from guilt aversion and whether guilt aversion toward the recipient is stronger than toward the donor. Testing the predictions of the model in a laboratory experiment, we show that the proportion of guilt-averse intermediaries is the same irrespective of the direction of the guilt. However, structural estimates indicate that the effect of guilt on behaviour is higher when the guilt is directed toward the recipient.

JEL code: C91
Keywords: Embezzlement, Dishonesty, Guilt Aversion, Psychological Game Theory, Experiment

[a] University of Lille, Lille Economics Management, Faculté des Sciences Economiques et Sociales, Cité Scientifique, 59655 Villeneuve d'Ascq Cedex, France. E-mail: giuseppe.attanasi@univ-lille1.fr
[b] Univ Lyon, University Lyon 2, GATE UMR 5824, 93 Chemin des Mouilles, F-69130, Ecully, France. E-mail: rimbaud@gate.cnrs.fr
[c] Univ Lyon, CNRS, GATE UMR 5824, 93 Chemin des Mouilles, F-69130, Ecully, France. IZA, Bonn, Germany. E-mail: villeval@gate.cnrs.fr

Acknowledgements: We are grateful to O. Armantier, P. Battigalli, C. Bellemare, M. Dufwenberg, A. Festré, D. Gambetta, A. Guido, J. Hamman, N. Hanaki, N. Köbis, D. Nosenzo, A. Rustichini, and R. Slonim for very valuable feedback. We thank also participants at the BEERS seminar (Lyon), the Bocconi student seminar (Milan), the 1st Summer School on Psychological Game Theory (Norwich), the 2nd workshop on Psychological Game Theory (Norwich), the CREED Lunch seminar (Amsterdam) and the LEEN seminar (Nice) for very useful comments. Financial support from the FELIS program of the French National Agency for Research (ANR-14-CE28-0010-01) is gratefully acknowledged. This research has also benefited from the support of the Van Gogh PHC program and of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-007) operated by the French National Research Agency (ANR).

1 Introduction

Embezzlement, defined as the misappropriation of assets by individuals to whom they were entrusted in order to monopolize or to steal them, is widespread all around the world. It can be observed in any situation in which the providers of resources need intermediaries to transfer these resources to the final recipients. The problem is particularly important in developing countries, in various domains such as health,[1] education,[2] or humanitarian aid,[3] where the final recipients seldom receive the totality of governmental or international aid transfers they are entitled to. In these countries, governmental and non-governmental organizations must rely on local intermediaries and cannot easily verify which amount has eventually been transferred to the entitled recipients. It is a major concern since embezzlement is detrimental to economic development (Olken and Pande, 2012). Once the costs of embezzlement are accounted for, some programs may even be inequality enhancing (*e.g.*, Reinikka and Svensson, 2014) or no longer cost-effective (*e.g.*, Ferraz *et al.*, 2012).

Exploring the determinants of the intermediaries' behavior in this context is crucial to design policies that are better able to deter this type of fraud. So far, the literature has focused mainly on the monetary determinants of embezzlement in line with the traditional

[1] In Cambodia, in 2005, 5% to 10% of the Health budget "disappeared" before it was paid from the Ministry of Finance to the Ministry of Health (Gaitonde *et al.*, 2016). In Ghana, in 2000, the Public Expenditure Tracking Survey revealed that 80% of non-salary funds did not reach health facilities (Canagarajah and Ye, 2001).

[2] For the period 1991-1995, Ugandan schools received on average 13% of the governmental transfers they were entitled to (Reinikka and Svensson, 2004). Ferraz *et al.* (2012) report at least 26 cases of embezzlement of decentralized educational grants in Brazil. In the municipality of Placas, between 2003 and 2004, US$ 1.25 millions of decentralized public educational grants could not be accounted for. Embezzlement even extended to stealing teacher's wages: in Itabuna, in May 2009, approximately 90% of municipal school teachers, received less than half of their monthly salaries.

[3] In 2013, the head of the governmental High Relief Committee was arrested for the misappropriation of US$ 10 million earmarked for the aid of refugees in Lebanon. In India, in 2005, government and bank officials were charged with embezzling US$ 2.5 millions in state funds designated for flood relief efforts.

economics-of-crime approach (*e.g.*, Becker and Stigler, 1974; Fan*et al.*, 2009). This has led to explore policy instruments such as increasing the officers' wages or the audit intensity (*e.g.*, Di Tella and Schargrodsky, 2003; Olken, 2007; Barr *et al.*, 2009). But like other forms of dishonesty (*e.g.*, Mazar *et al.*, 2008; Fischbacher and Föllmi-Heusi, 2013; Abeler *et al.*, 2018), embezzlement and corruption also entail moral costs (*e.g.*, Abbink and Serra, 2012; Drugov *et al.*, 2014) that may not be sensitive to these interventions.

In this paper we focus on the psychological cost of embezzling, by studying the intermediary's willingness to avoid the anticipated negative valence associated with guilt from embezzlement. Guilt aversion implies that an agent suffers a cost, *i.e.*, feels guilty, if he lets down others' expectations (Tangney and Fischer, 1995). Theoretically, we rely on the modeling of guilt aversion as a belief-dependent motivation by Charness and Dufwenberg (2006) and Battigalli and Dufwenberg (BD, hereafter) (2007) in the framework of psychological game theory introduced by Geanakoplos *et al.* (1989) and further developed by Battigalli and Dufwenberg (2009). This theory departs from traditional game theory in assuming that players' utilities do not only depend on their actions but also on their beliefs about choices, beliefs, or information. In particular, the psychological utility of a guilt-averse player depends on his second-order beliefs, *i.e.*, his beliefs about the other players' beliefs about his own decision.

Our aim is to identify the existence and the direction of guilt aversion and its impact on the behavior of intermediaries who have the opportunity to embezzle the donations made by donors to recipients. While embezzling donations increases the intermediaries' monetary payoff, it decreases both the utility of donors who care about the recipients' wellbeing and the utility of the recipients who expect to receive a donation. As a result, intermediaries who embezzle may feel guilty both toward the donor and toward the recipient. Indirect evidence of intermediaries' guilt aversion can be found in previous

experimental studies of embezzlement. Chlaß *et al.* (2015) found that the more donors believe that the donation will be transferred, the more they donate, and the more intermediaries believe that donors have donated, the more they transfer. [4] This two-fold result is consistent with our model of guilt aversion which predicts that the more donors believe the donation will be transferred, the more intermediaries transfer. Meanwhile, conducting a lab-in-the-field experiment in Tanzania, Di Falco *et al.* (2016) have shown that intermediaries at the beginning of longer chains embezzle less than intermediaries in short chains, anticipating that other intermediaries after them may embezzle as well. Feeling guilty from letting down the recipients' expectation might explain the intermediaries' behaviour in their experiment. Interestingly, the idea that guilt aversion is a relevant motivation in the context of corruption was also developed by Balafoutas (2011). The author provides a dynamic model where a guilt-averse official may feel guilty toward a citizen if he accepts the bribe offered by a lobby. Our contribution goes one step further by providing an experimental test of guilt-aversion as a determinant of corrupt behaviour.

Specifically, we model the intermediary's guilt aversion both toward the donor and the recipient and further test whether one direction matters more than the other. By doing so, we contribute both to the theoretical extension of BD (2007) model of guilt aversion and to the empirical identification of determinants of the embezzlement decision. If we find a difference between guilt aversion *vis-à-vis* the donor and *vis-à-vis* the recipient, it may also suggest different policy interventions.

We apply BD (2007) model of simple guilt to a novel three-player game – the Embezzlement Mini-Game. In this game, a donor sends a donation to a recipient but it has to be transferred by an intermediary who can embezzle a fraction of the donation. The

[4] We reckon that the intermediaries' behaviour may be driven by the fact that reporting beliefs after deciding allows the more dishonest intermediaries to justify their deed. We control for this bias by inducing second-order beliefs rather than relying on reported beliefs.

intermediary may be affected by two sources of guilt aversion: donor-guilt aversion and recipient-guilt aversion. The donor entrusts him with a donation and forms expectations on his willingness to transfer the whole donation to the recipient.[5] The recipient also has expectations on how much he will receive. Depending on his decision, the intermediary can fulfill (or not) the other two players' expectations. Hence, he may be guilt-averse both toward the donor and toward the recipient.

For recipient-guilt, we rely on BD's definition of guilt toward another player as the disutility from letting down the other player's expectations about *his own* payoff. For donor-guilt, we extend this definition of guilt. Rather than not letting down the donor's expectations about his own payoff – which is not affected by the intermediary's decision to embezzle –, a donor-guilt averse intermediary dislikes letting down the donor's expectations about *another player*'s monetary payoff, *i.e.*, the recipient. Thus, in this case, the psychological utility of the guilt-averse player (the intermediary) directly depends on his beliefs about another player's belief (the donor) on a*third* player's actual monetary payoff (the recipient).[6] To the best of our knowledge, the latter modelling choice is new in the psy-games literature.[7]

Our theoretical analysis builds on the incomplete-information environment with role-dependent guilt of Attanasi *et al.* (2016). We assume that among the two active players in the embezzlement game only the intermediary can feel guilty and we allow for bi-dimensional guilt. However, unlike in Attanasi *et al.* (2016), we use rationalizability

[5] "Ensuring that a reasonable proportion of donations get to end cause" is the most important quality of a charity according to survey conducted in Wales and England (Ipsos MORI, 2010).

[6] Indeed, this extension relies on the auxiliary assumption that the recipient's payoff enters the utility of an altruistic donor and the intermediaries can infer that the donor is altruistic in this sense.

[7] In Balafoutas (2011), an extension of his model also allows for the official to feel guilty toward both the citizen and the lobby. However, both citizen-guilt and lobby-guilt are coherent with BD's original of model since the official can affect the payoff of both the citizen and the lobby.

rather than Bayesian equilibrium as a solution concept.[8] We apply the forward-induction rationalizability analysis elaborated by Attanasi *et al.* (2013) for Trust Mini-games. However, our analysis is complicated by the fact that the donor's sensitivity to altruism is unknown to the intermediary whereas, in Attanasi *et al.* (2013), the trustor is commonly known to be selfish.

We tested the predictions of our model by implementing our Embezzlement Mini-Game in a laboratory experiment that allows us to measure directly the role of second-order beliefs on the intermediaries' decision to embezzle donations, adapting the belief-dependent menu method of Khalmetski *et al.* (2015). Within-subjects, we manipulated the percentage of the donation that could be embezzled: 80% in the High condition and 60% in the Low condition. Between-subjects, we manipulated the information given to the intermediaries before they made their decision. In the Donor treatment, intermediaries decided whether transferring or not the whole donation for each possible first-order belief of the donor on their decision. In the Recipient treatment, they made a decision for each possible first-order belief of the recipient on their decision. We can therefore compare the intermediaries' donor-guilt aversion and recipient-guilt aversion. We also elicited the individual sensitivity to guilt by using the Guilt and Shame Proneness (GASP) questionnaire of Cohen *et al.* (2011).

We have two main findings. First, we show that, on average, 25% of the intermediaries are guilt-averse, *i.e.*, their decision to embezzle is influenced by others' expectations, and this holds regardless of the direction of the guilt. Second, structural estimates indicate that guilt sensitivity is higher toward the recipients' expectations than toward the donors' expectations. Taken together, these results suggest that to discourage embezzlement,

[8]Using rationalizability is motivated by the fact that a standard equilibrium analysis has no compelling foundation for games played one-shot (like ours) and in general in experiments on other-regarding preferences.

policy interventions can try to increase the anticipation of guilt among intermediaries by disseminating information on others' expectations toward them, both the expectations of the donors and those of the recipients.

The remainder of this paper is organized as follows. Section 2 introduces the theoretical model and its behavioral predictions. We describe the experimental design in Section 3.1. Section 4 presents the experimental results. Section 5 discusses and concludes.

2 Theoretical Model and Behavioral Hypotheses

2.1 The Embezzlement Mini-Game(s)

The Embezzlement Mini-Game involves three players: a donor, an intermediary and a recipient (see Figure 1). Players' material payoffs in Figure 1 are shown according to such order.

The three players receive an initial endowment: 150 ECU (Experimental Currency Units) for the donor, 80 ECU for the intermediary, and 10 ECU for the recipient (with 10 ECU = €1.2 in the experiment). Thus, the intermediary's endowment is median between the donor's and the recipient's endowments.[9]

The donor can keep his endowment (in which case the game ends and each player earns his endowment) or give 25 ECU to the recipient. However, the donation cannot be given *directly* to the recipient, it has to be transferred through the intermediary. The intermediary has to decide whether to transfer the entirety of the donation to the recipient or to embezzle a fraction f of the 25 ECU and transfer $(1–f)$ to the recipient. The recipient

[9]The intermediary can be seen as the middleman in a network linking a NGO or a Governmental Agency to villagers. Unlike in a consecutive three-person dictator game (Bahr and Requate, 2013), the different initial endowments underline the different status of each player.

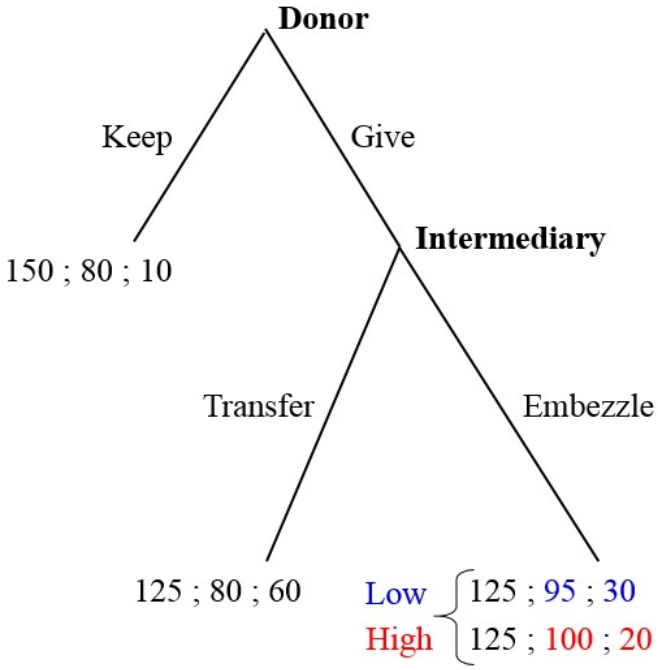

Figure 1: The Embezzlement Mini-Game(s)

receives twice the amount actually transferred. Thus, embezzlement involves an efficiency loss.[10]

Using a within-subject design, the Mini-Game is played under two conditions, each one allowing the intermediary to embezzle a different fraction of the donation: in the Low condition, $f = 0.6$, and in the High condition, $f = 0.8$. Therefore, the two Mini-Games only differ for the set of possible actions of the intermediary (respectively, $f \in \{0, 0.6\}$ and $f \in \{0, 0.8\}$), both being a strict subset of the possible actions in the original game, where $f \in [0, 1]$.

Figure 1 also shows two features of the final payoff distributions under each of these two conditions. First, no decision can lead to the equalization of payoffs between two or

[10]This feature (also used in Boly *et al.*, 2016) captures a negative externality associated with embezzlement (see Ferraz *et al.*, 2012, for an illustration in the domain of education in Brazil). The presence of a negative externality also reinforces the immoral image of embezzlement.

three players. Hence, no payoff distribution should be more salient than others. Second, the ranking of payoffs cannot be affected by the players' decisions. By doing so, we limit social comparison motives.

2.2 Utility Functions

2.2.1 Assumptions about Social Preferences

We assume that the donor can be motivated by *distributional preferences* – altruism (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002), and the intermediary can be motivated by *belief-dependent preferences* – guilt aversion *à la* BD (2007).

We assume that *the donor's utility increases with the recipient's payoff while it is unaffected by the intermediary's payoff*. This is a simplifying assumption but it is broadly consistent with other models of distributional preferences. Donors would also prefer that their donation increases the recipient's payoff rather the intermediary's if they were inequality-averse (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), since the recipient is the most disadvantaged player; and if they were concerned with efficiency (as in Charness and Rabin, 2002) since social welfare is maximum if the entire donation is transferred to the recipient. Our experimental design allows us to test this assumption: we elicit the donor's first-order belief that the intermediary will choose *Transfer* after *Give*. If the donor's utility increases with the recipient's payoff about the recipient, we should find that the probability of giving increases in the donor's first-order belief about *Transfer* after *Give*.

We also assume that *the intermediary can feel guilty toward both the recipient and the donor*: he might want not to disappoint the recipient's and/or the donor's expectations

about the amount he will actually transfer conditionally on the donor's giving. In our model, we analyze the impact of each guilt sensitivity (toward the donor and toward the recipient) separately, since we used a between-subject design to elicit the intermediary's belief-dependent strategy conditional on either the donor's (Donor treatment) or the recipient's (Recipient treatment) first-order beliefs (see Section 3.1). Therefore, we make the auxiliary assumption that *in the Donor (Recipient) treatment, guilt toward the donor (the recipient) prevails over guilt toward the recipient (the donor)* as belief-dependent motivation of the intermediary's behavior.

We analyze the Embezzlement Game under *incomplete information of players' social preferences*, *i.e.*, we assume that the donor's sensitivity to altrusim toward the recipient, and the intermediary's sensitivities to guilt toward either the donor or the recipient, are not common knowledge. Although this yields weaker predictions than under complete information, we believe that this condition better fits the environment of our experiment, where, as in the overwhelming majority of other laboratory studies on belief-dependent preferences, sensitivities to such preferences are not disclosed within the matched subjects in the game (for an exception, see Attanasi *et al.* (2013, 2018)).

2.2.2 The Donor's Utility Function

The donor's utility function is composed of two parts (Eq. 1).

$M_D(s_D)$ is the donor's material payoff. It only depends on his strategy $s_D \in \{Keep, Give\}$. In the case he keeps, his material payoff coincides with his endowment. In the case he gives, the endowment is reduced by the amount given.

$A_D(\gamma_D, s_I)$ represents the donor's altruism toward the recipient, *i.e.*, his preferences toward an increase in the recipient's material payoff. It is the product of two terms: (i) $\gamma_D \geq 0$, the donor's willingness-to-pay to increase the recipient's payoff by 2 (since

every ECU transferred to the recipient is doubled): it is an exogenous parameter that describes the donor's type in terms of *sensitivity to altruism* toward the recipient; (ii) $r(s_I)$, how much the recipient eventually receives: it depends on s_I, the intermediary's strategy conditional on the donor choosing *Give*.

In fact, the donor's altruism toward the recipient enters his utility function only if he chooses *Give*, *i.e.*, $s_D = G$. In that case, the donor's actual altruism toward the recipient depends on the intermediary's action. We denote with α_{DI} the donor's first-order belief that the intermediary chooses *Transfer*, conditional on the donor choosing *Give*. With this, the utility (Eq. 1) and expected utility (Eq. 2) of the donor go as follows:

$$U_D(s_D, s_I, \gamma_D) = M_D(s_D) + \mathbb{1}_{s_D=G} \cdot A_D(\gamma_D, s_I), \text{ with } A_D(\gamma_D, s_I) = \gamma_D \cdot r(s_I) \quad (1)$$

$$\mathbb{E}_D[U_D(s_D, \alpha_{DI}, \gamma_D)] = M_D(s_D) + \mathbb{1}_{s_D=G} \cdot \gamma_D \cdot (\alpha_{DI} \cdot r(T) + (1 - \alpha_{DI}) \cdot r(E)) \quad (2)$$

We assume that the donor's sensitivity to altruism, γ_D, can take two values. Either the donor is *selfish*, $\gamma_D = 0$, or he is *altruistic*, $\bar{\gamma}_D > 1/2$. The motivation for the specific threshold of $1/2$ for an altruistic donor will be provided in Section 2.3.1, when discussing the rational behavior of an altruistic donor.

2.2.3 The Recipient's Utility Function

The recipient's utility function coincides with his material payoff (Eq. 3). The material payoff is made of two parts: (i) M_R is the recipient's endowment, that he holds independently of the donor's and intermediary's choices ; (ii) $M_R(s_D, s_I)$ is the part that depends on the donor's and the intermediary's strategies. It enters the recipient's utility function only if the donor chooses *Give*, *i.e.*, $s_D = G$. In that case, it depends on the intermediary's strategy: $M_R(s_D, s_I) = r(s_I)$, *i.e.*, the amount actually received, that we also used to measure the donor's altruism in Eq. (2). With this, the utility function of the recipient

is:

$$U_R(s_D, s_I) = M_R + \mathbb{1}_{s_D=G} \cdot r(s_I) \tag{3}$$

We are interested in the recipient's beliefs only in terms of their psychological impact on the intermediary's choice. Thus, in the experiment we only elicited α_{RI}, namely the recipient's first-order belief that the intermediary chooses *Transfer*, conditional on the donor choosing *Give* (and not the recipient's belief about the donor choosing *Give*). Hence, we only focus on the recipient's expected utility after the donor has made his choice:

$$\mathbb{E}_R[U_R(s_D, \alpha_{RI})] = M_R + \mathbb{1}_{s_D=G} \cdot (\alpha_{RI} \cdot r(T) + (1 - \alpha_{RI}) \cdot r(E)) \tag{4}$$

2.2.4 The Intermediary's Utility Function

The intermediary's utility function is composed of three parts (Eq. 5). The first part corresponds to his material payoff, $M_I(s_D, s_I)$. It depends on both the donor's and his own strategy. If $s_D = G$, it depends on his own action, otherwise it coincides with his endowment.

The second part, G_{ID}, corresponds to the intermediary's feeling of guilt toward the donor. It is the product of two terms: (i) $\theta_{ID} \geq 0$, the guilt sensitivity toward the donor; (ii) the difference, if positive, between the donor's expected transfer $\mathbb{E}_D[r(\alpha_{DI})]$ and the realized transfer $r(s_I)$ to the recipient. Thus, (i) is an exogenous parameter that describes the intermediary's type, while (ii) depends both on the intermediary's strategy, and on the donor's conditional first-order belief about this strategy. If $\mathbb{E}_D[r(\alpha_{DI})] > r(s_I)$, the intermediary feels guilty from letting down the donor's expectations on the amount

transferred by the recipient.

The third part, G_{IR}, captures the intermediary's feeling of guilt toward the recipient. It is formed similarly as for G_{ID} as the product of (i) $\theta_{IR} \geq 0$, the guilt sensitivity toward the recipient, and (ii) the difference, if positive, between the recipient's expected transfer $\mathbb{E}_R[r(\alpha_{RI})]$ and the realized transfer $r(s_I)$.

Thus, the intermediary's psychological utility is defined as:

$$\begin{aligned} U_I(s_D, s_I, \theta_{ID}, \theta_{IR}, \alpha_{DI}, \alpha_{RI}) &= M_I(s_D, s_I) - G_{ID}(s_I, \theta_{ID}, \alpha_{DI}) - G_{IR}(s_I, \theta_{IR}, \alpha_{RI}) \\ \text{where } G_{ID} &= \theta_{ID} \cdot max\{0, \mathbb{E}_D[r(\alpha_{DI}] - r(s_I)\} \\ \text{and } G_{IR} &= \theta_{IR} \cdot max\{0, \mathbb{E}_R[r(\alpha_{RI})] - r(s_I)\} \end{aligned} \tag{5}$$

If $s_D = K$ (the donor keeps), then $\mathbb{E}_D[r(\alpha_{DI})] = \mathbb{E}_R[r(\alpha_{RI})] = 0$ and the intermediary cannot feel guilty. It does not feel guilty also if $(s_D, s_I) = (G, T)$, *i.e.*, the donor gives and the intermediary transfers the whole donation to the recipient.

We assume that there are only two psychological types for the intermediary, given the specific source of guilt. Thus, with regard to the donor, we assume that the intermediary can be either *donor-selfish*, $\underline{\theta}_{ID} = 0$, or *donor-guilt-averse*, $\bar{\theta}_{ID} > 1/2$. As for the recipient, he can be either *recipient-selfish*, $\underline{\theta}_{IR} = 0$, or *recipient-guilt-averse*, $\bar{\theta}_{IR} > 1/2$. The motivation for the specific threshold of 1/2 for a guilt-averse intermediary will be provided in Section 2.3.2, when discussing the rationalizability of a guilt-averse intermediary's behavior.

Two clarifications are in order about Equation (5). First, in Equation (5) we measure the intermediary's guilt aversion through the co-players' first-order rather than with the intermediary's second-order beliefs about *Transfer* choices. Given our design (see section 3.1), when the intermediary takes his decision, the donor's and the recipient's beliefs

are observed, hence the intermediary's second-order belief corresponds to the donor's or recipient's first-order belief.[11]

Second, as anticipated, although donor-guilt aversion and recipient-guilt aversion appear simultaneously in Equation (5), we will analyze the intermediary's behavior separately for G_{ID} and G_{IR}, under the assumption that $\theta_{IR} = 0$ ($\theta_{ID} = 0$) in the Donor treatment (Recipient treatment).[12]

2.3 Rationalizability with Forward Induction

We provide a rationalizability analysis of the Embezzlement Mini-Game(s) with incomplete information based on forward-induction reasoning (cf. BD, 2009, Section 5; Battigalli *et al.*, 2017). Under incomplete information, both the psychological type of the donor (γ_D) and the psychological type of the intermediary (θ_{ID} or θ_{IR}) are not common knowledge, while the recipient is commonly known to be selfish.[13]

The first two steps of our analysis are based on the following assumptions:

Step 1. **Rationality**: each player is rational, *i.e.*, a subjective expected utility maximizer.

Step 2. **Strong belief in rationality:** each player is certain of the rationality of the co-player as long as such rationality is not contradicted by observed behavior.

[11] BD (2009) have shown the equivalence of the analysis in terms of other's first-order or own second-order beliefs in dynamic psychological games.

[12] This is without loss of generality, especially under our rationalizability approach: we are not interested in equilibrium thresholds of θ_{ID} and θ_{IR}, but only on the correlation between these guilt sensitivities and the intermediary's behavior, and both guilt components push the intermediary in the direction of *Transfer* over *Embezzle* for positive co-player's beliefs. Moreover, analyzing the two cases separately allows us to provide sharper predictions in each of the two treatments, which is the main motivation behind our between-subject design.

[13] See Appendix A for a presentation of the Harsanyi-type structure behind our incomplete-information game. There we also discuss the main methodological advantages of the two-step rationalizability procedure that we use to elaborate behavioral predictions in the incomplete-information game.

The second assumption is the basic **forward-induction reasoning** (see Battigalli and Siniscalchi, 2002; BD, 2009).

2.3.1 Rationality of the Donor

We perform the first step of rationalizability (best-reply analysis) of the donor's behavior in the two conditions of our embezzlement game, namely L (Low) and H (High). Relying on Eq. (2), we define the donor's **Willingness-to-Give function (WG)** as the difference between his expected utility from *Give* and his (certain) utility from *Keep* (Eq. 6).

$$\begin{aligned} WG(\alpha_{DI}, \gamma_D) &= \mathbb{E}_D[U_D(G, \alpha_{DI}, \gamma_D] - U_D(K, \gamma_D) \\ &= M_D(G) + \gamma_D \cdot (\alpha_{DI} \cdot r(T) + (1 - \alpha_{DI}) \cdot r(E)) - M_D(K) \end{aligned} \tag{6}$$

Recall that we have assumed that the donor's set of utility types is made of only two elements, *i.e.*, $\gamma_D = \{0, \bar{\gamma}_D\}$, with $\bar{\gamma}_D > 1/2$. The rationality of a selfish donor leads to a negative Willingness-to-Give function in both the Low and the High conditions:

$$WG^k(\alpha_{DI}, 0) = -25, \text{ with } k = L, H$$

The rationality of an altruistic donor leads to the following *Give* thresholds for $\bar{\gamma}_D$:

$$WG^L(\alpha_{DI}, \gamma_D) = -25 + \bar{\gamma}_D \cdot (\alpha_{DI} \cdot 30 + 20) > 0 \qquad \Longleftrightarrow \qquad \bar{\gamma}_D > \frac{5}{6\alpha_{DI} + 4} \tag{7}$$

$$WG^H(\alpha_{DI}, \gamma_D) = -25 + \bar{\gamma}_D \cdot (\alpha_{DI} \cdot 40 + 10) > 0 \qquad \Longleftrightarrow \qquad \bar{\gamma}_D > \frac{5}{8\alpha_{DI} + 2} \tag{8}$$

These inequalities justify our assumption of $\bar{\gamma}_D > 1/2$ for an altruistic donor. Consider the situation where an altruistic donor is paired with an intermediary believed to *Transfer*

(*i.e.*, $\alpha_{DI} = 1$). On the one hand, if $\bar{\gamma}_D = 1/2$, an altruistic donor would be indifferent between *Give* or *Keep* and, if $\bar{\gamma}_D < 1/2$, an altruistic donor would behave as a selfish one. On the other hand, if $\bar{\gamma}_D > 1/2$, an altruistic donor would *Give* absent the intermediary, *i.e.*, he would be altruistic in the strategically equivalent Dictator game.

Recall that in our experiment the set of possible first-order beliefs of the donor is discrete: $\alpha_{DI} \in \{0, 1/3, 2/3, 1\}$. Substituting these values into Eqs. (7) and (8), we find the set of 'belief – type' pairs consistent with a rational donor choosing *Give* in the Low and High conditions, respectively (Step 1):

$$\begin{aligned} R_D^{1,G|L} &= \{(\alpha_{DI}, \bar{\gamma}_D) : (\alpha_{DI} = 0, \bar{\gamma}_D > 5/4), (\alpha_{DI} = 1/3, \bar{\gamma}_D > 1), \\ &\quad (\alpha_{DI} = 2/3, \bar{\gamma}_D > 5/8), (\alpha_{DI} = 1, \bar{\gamma}_D > 1/2)\} \\ R_D^{1,G|H} &= \{(\alpha_{DI}, \bar{\gamma}_D) : (\alpha_{DI} = 0, \bar{\gamma}_D > 5/2), (\alpha_{DI} = 1/3, \bar{\gamma}_D > 15/14), \\ &\quad (\alpha_{DI} = 2/3, \bar{\gamma}_D > 15/22), (\alpha_{DI} = 1, \bar{\gamma}_D > 1/2)\} \end{aligned} \tag{9}$$

Now, fix the donor's type $\bar{\gamma}_D$ and consider the (rationality) requirement that action *Give* maximizes the donor's expected utility in the Low condition ($R_D^{1,G|L}$). If $\bar{\gamma}_D \in (1/2, 5/8)$, this donor's type will *Give* only for $\alpha_{DI} = 1$; if $\bar{\gamma}_D \in (5/8, 1)$, he will also *Give* for $\alpha_{DI} = 2/3$; if $\bar{\gamma}_D \in (1, 5/4)$, he will also *Give* for $\alpha_{DI} = 1/3$; if $\bar{\gamma}_D \in (5/4, +\infty)$, he will *Give* for every α_{DI}. Therefore, independently from $\bar{\gamma}_D$, $\partial WG^k/\partial \alpha_{DI} \geq 0$ for $k = L, H$. Thus, considering heterogeneity in donors' types (sensitivity to altruism), we expect the frequency of *Give* choices to increase with α_{DI}. A similar prediction can be elaborated for the High condition. All this leads to our first behavioral prediction about the donor's belief-dependent behavior.

H.D1 [Choice-Belief Correlation]: The frequency of *Give* choices by altruistic donors increases in their first-order belief about *Transfer*.

Note that is H.D1 is not verified, it contradicts our assumption that the donor's utility increases with the recipient's payoff while it is unaffected by the intermediary's payoff.

Furthermore, consider the comparison between the Willingness-to-Give function of an altruistic donor in the two conditions:

$$WG^L(\alpha_{DI}, \bar{\gamma}_D) - WG^H(\alpha_{DI}, \bar{\gamma}_D) = 5 + \bar{\gamma}_D \cdot 10 \cdot \alpha_{DI}$$

We see that, given that $\bar{\gamma}_D > 1/2$, then $WG^L - WG^H > 0$ for $\alpha_{DI} < 1$. Note that we can assume that the distribution of the donors' types is the same across conditions since conditions are manipulated within-subjects. Therefore, we can construct the set of 'belief – type' pairs consistent with a rational donor choosing *Keep* in the Low condition and *Give* in the High condition, $(R_D^{1,K|L} \cap R_D^{1,G|H}) = \varnothing$, and the set consistent with a rational donor choosing *Give* in the Low condition and *Keep* in the High condition, $(R_D^{1,G|L} \cap R_D^{1,K|H}) = \{(\alpha_{DI} = 0, 5/4 < \bar{\gamma}_D < 5/2), (\alpha_{DI} = 1/3, 1 < \bar{\gamma}_D < 15/14), (\alpha_{DI} = 2/3, 5/8 < \bar{\gamma}_D < 15/22)\}$. The following behavioral hypothesis summarizes this prediction.

H.D2 [High *vs.* Low Condition on Choice]: The frequency of *Give* choices by altruistic donors is higher in the Low than in the High condition for each donor's first-order belief lower than one.

Note that H.D1 and H.D2 should hold in both the Donor and the Recipient treatment.

2.3.2 Rationality and Forward Induction of the Intermediary

We analyze the intermediary's behavior in the Donor treatment and the Recipient treatment separately, assuming $\theta_{IR} = 0$ in the former and $\theta_{ID} = 0$ in the latter. We begin with the **Donor treatment**, since it is the more interesting due to the above mentioned

extension of BD (2007) in the definition of donor-guilt aversion.

Relying on Eq. (5), we define the intermediary's **Willingness-to-Transfer function** (**WT**) – conditional on Give[14] – as the difference between his utility when he Transfers (thereby experiencing no guilt) and his expected utility when he Embezzles a fraction of the donation (Eq. 10).

$$\begin{aligned} WT(\alpha_{DI}, \theta_{ID}|G) &= \mathbb{E}_I[U_I(\alpha_{DI}, \theta_{ID})|G, T] - \mathbb{E}_I[U_I(\alpha_{DI}, \theta_{ID})|G, E] \qquad (10) \\ &= M_I(G, T) - M_I(G, E) + \theta_{ID}(\alpha_{DI} \cdot r(T) + (1 - \alpha_{DI}) \cdot r(E)) - r(E)) \end{aligned}$$

The analysis in each condition depends on the shape of $WT(\alpha_{DI}, \theta_{ID}|G)$ implied by the intermediary's psychological type θ_{ID}. Rationality implies that the intermediary of type θ_{ID} chooses the dominant action when it exists. This gives the step-1 prediction set R_I^1.

In the second step, it is assumed that the intermediary strongly believes in the donor's rationality. Therefore, after having observed *Give*, the intermediary can infer the altruistic type of the donor, *i.e.*, $\gamma_D = \bar{\gamma}_D$ independently from the donor's first-order belief of *Transfer* (*Give* is never a best reply for a selfish donor); formally, $\mathbb{P}_I\left(R_D^1|G\right) = \mathbb{P}_I\left(\tilde{\gamma}_D = \bar{\gamma}_D|G\right) = 1$. Thus, the intermediary believes that the donor's Willingness-to-Give function is represented by (Eq. 10) *i.e.*), that the donor's utility increases with $r(s_I)$. With this, a 'choice – type' pair (s_I, θ_{ID}) is consistent with the intermediary's rationality and strong belief in the donor's rationality if and only if there is some α_{DI}

[14]The intermediary's decision is made under the strategy method (*i.e.*, both when the donor has chosen *Keep* and when he has chosen *Give*). Here we assume that the intermediary best-responds *as if* he had truly observed the donor's move. This holds by standard expected-utility maximization, except for the cases where the intermediary is certain that the donor has chosen *Keep*. Thus, we need the additional assumption that the intermediary has a belief conditional on *Give*, even if he is certain of *Keep*.

such that $WT(\alpha_{DI}, \theta_{ID}|G) > 0$. Let

$$R_I^{2,T} = \left\{ (s_I, \theta_{ID}) : \max_{\alpha_{DI} \in \{0, \frac{1}{3}, \frac{2}{3}, 1\}} WT(\alpha_{DI}, \theta_{ID}|G) > 0, s_I = T \right\}$$

and

$$R_I^{2,E} = \left\{ (s_I, \theta_{ID}) : \max_{\alpha_{DI} \in \{0, \frac{1}{3}, \frac{2}{3}, 1\}} WT(\alpha_{DI}, \theta_{ID}|G) < 0, s_I = E \right\}$$

then the second step of rationalizability leads to $R_I^2 = R_I^{2,T} \cup R_I^{2,E}$. *Transfer* (respectively, *Embezzle*) is forward-induction dominant for θ_{DI} if and only if $WT(\alpha_{DI}, \theta_{ID}|G) > 0$ (respectively $WT(\alpha_{DI}, \theta_{ID}|G) < 0$) *for every* $\alpha_{DI} \in \{0, \frac{1}{3}, \frac{2}{3}, 1\}$.

For a donor-selfish intermediary ($\theta_{ID} = 0$), the Willingness-to-Transfer function (Eq. 10) is negative, independently from both the first-order belief of *Transfer* and the condition, *i.e.*, a donor-selfish intermediary will always choose *Embezzle* in both the Low and the High condition.

For a donor-guilt-averse intermediary ($\theta_{ID} = \bar{\theta}_{ID} > 1/2$), the Willingness-to-Transfer function is positive, independently from the condition, if $\alpha_{ID} = 1$ i.e. a donor-guilt-averse intermediary will choose to *Transfer* if (he is certain that) the donor is certain of him choosing *Transfer*. This motivates our assumed threshold of $\bar{\theta}_{ID} > 1/2$ for an intermediary to be guilt-averse.

The donor choosing *Give*, thereby revealing his type $\bar{\gamma}_D > 1/2$, induces a restriction on the intermediary's (second-order) belief of α_{DI}. Consider $R_D^{1,G|L}$, the set of donors' 'belief – type' pairs consistent with rationality in the Low condition (Eq. 9). If, *e.g.*, $\bar{\gamma}_D = 3/4$, then the intermediary believes that $\alpha_{DI} = \{2/3, 1\}$, *i.e.*, he believes that the donor has chosen *Give* believing that at least two over the three intermediaries who can be randomly matched with him would choose *Transfer*. This leads to $\min_{\alpha_{DI} \in \{2/3,1\}} WT^L(\alpha_{DI}, \theta_{ID}|G) > 0$ iff $\bar{\theta}_{ID} > 3/4$, *i.e.*, *Transfer* is the dominant action

(and hence rationalizable) iff $\theta_{ID} = \bar{\theta}_{ID} > 3/4$. If instead $\bar{\gamma}_D = 9/8$ (the donor is more altruistic), then the intermediary believes that $\alpha_{DI} = \{1/3, 2/3, 1\}$ (the donor might have chosen *Give* also for 1/3), and $\min_{\alpha_{DI} \in \{1/3,2/3,1\}} WT^L(\alpha_{DI}, \theta_{ID}|G) > 0$ iff $\bar{\theta}_{ID} > 3/2$. Therefore, $R_I^{2,T}$ shrinks if the smallest second-order belief of *Transfer* consistent with (rationality and) forward induction decreases. This, coupled with the heterogeneity of psychological types (the sensitivity to donor-guilt) in the population of intermediaries, leads to conclude that the frequency of *Transfer* choices should increase with the intermediary's assessment of α_{DI}. The same result holds in the High condition. Indeed, given $\theta_{ID} > 0$, $\partial WT^k/\partial \alpha_{DI} \geq 0$ for $k = L, H$. The following behavioral hypothesis summarizes our first prediction on the intermediary's belief-dependent behavior.

H.I1 [Choice-Belief Correlation]: The frequency of *Transfer* choices by donor-guilt-averse intermediaries increases in their second-order beliefs about *Transfer*.

Furthermore, for any α_{ID} and $\bar{\theta}_{ID}$, it is $WT^H(\alpha_{DI}, \bar{\theta}_{ID}|G) = (4/3) \cdot WT^L(\alpha_{DI}, \bar{\theta}_{ID}|G)$. We can assume that the distribution of the intermediaries' psychological types is the same across conditions since conditions are manipulated within-subjects. Therefore, the set of 'belief – choice' pairs consistent with a rational intermediary, who strongly believes in the donor's rationality, choosing *Transfer* is the same under both conditions ($R_I^{2,T|L} = R_I^{2,T|H}$) ; and the same holds for choosing *Embezzle* ($R_I^{2,E|L} = R_I^{2,E|H}$). The following behavioral hypothesis summarizes the prediction of insensitivity of guilt aversion to the change of condition.

H.I2 [High *vs.* Low condition on Choice]: The frequency of *Transfer* choices by intermediaries is the same in the Low and in the High conditions.

Finally, consider again the first example given above in the Low condition with an

induced donor's altruistic type $\bar{\gamma}_D = 3/4$, which leads the intermediary believe that $\alpha_{DI} = \{2/3, 1\}$. Consider two types of intermediaries, respectively $\bar{\theta}_{ID} = 3/5$ and $\bar{\theta}_{ID} = 1$. For type $\bar{\theta}_{ID} = 3/5$, neither *Transfer* nor *Embezzle* is forward-induction dominant, since $WT^L(\alpha_{DI}, 3/5|G) = -15 + 18 \cdot \alpha_{DI}$ (if $\alpha_{DI} = 2/3$ then $WT^L(2/3, 3/5|G) < 0$, while if $\alpha_{DI} = 1$ then $WT^L(1, 3/5|G) > 0$). Thus, his strategy depends on the precise value of his second-order belief, 2/3 or 1, which rationalizability does not pin down: if $\bar{\theta}_{ID}$ does not belong to either forward-induction dominance region $R_I^{2,T}$ or $R_I^{2,E}$, then both strategies can be justified by a second-order belief consistent with the assumption that the donor is rational. For type $\bar{\theta}_{ID} = 1$, *Transfer* is dominant, *i.e.*, he will choose *Transfer* independently of the second-order belief (which has been induced to be 2/3 or 1). Therefore, for the same induced beliefs, $\bar{\theta}_{ID} = 3/5$ might choose *Embezzle* while $\bar{\theta}_{ID} = 1$ will choose *Transfer*. This can also be interpreted: $\bar{\theta}_{ID} = 3/5$ has a minimum threshold for second-order beliefs that rationalizes the strategy *Transfer* that is higher than that of $\bar{\theta}_{ID} = 1$. All this also holds in the High condition with the same reasoning as for H.I2. The last behavioral hypothesis summarizes the two predictions above.

H.I3 [Choice-Guilt Correlation]: The higher the intermediary's donor-guilt-sensitivity, (i) the higher the frequency of *Transfer* choices, given his second-order belief about *Transfer*; (ii) the lower the minimum second-order belief sufficient to switch from *Embezzle* to *Transfer*.

Since we are not positing any specific assumption concerning the donor's exogenous beliefs about θ_{ID}, we cannot derive any further implication about the donor's behavior.[15] Because the third step does not refine the predictions for the donor, the incomplete-

[15]To see this, note that if in the Low condition a donor with $\bar{\gamma}_D = 1$ assigns more than 50% probability to $\theta_{ID} = \bar{\theta}_{ID} = 1$, then $\alpha_{DI} = 1$ and the best reply is *Give*; if, instead, he assigns more than 50% probability to $\theta_{ID} = \underline{\theta}_{ID} = 0$, then $\alpha_{DI} = 0$ and the best reply is *Keep*.

information rationalizability algorithm stops, *i.e.*, it gives the same predictions at each further step for each player.

Finally, note that the three behavioral hypotheses H.I1, H.I2, and H.I3 also hold in the **Recipient treatment**. The analysis in the Recipient treatment is easier since the recipient is commonly known to be selfish and the Willingness-to-Transfer function can be directly obtained without using forward induction to induce the recipient's type. In fact, at the first step the rationalizability procedure is the same. At the second step, the **Willingness to Transfer function** of a recipient-guilt averse intermediary (Eq. 11has the same functional form as $WT(\alpha_{DI}, \theta_{ID}|G)$ (Eq. 10).

$$\begin{aligned} WT(\alpha_{RI}, \theta_{IR}|G) &= \mathbb{E}_I[U_I(\alpha_{RI}, \theta_{IR})|G, T] - \mathbb{E}_I[U_I(\alpha_{RI}, \theta_{IR})|G, E] \\ &= M_I(G, T) - M_I(G, E) + \theta_{IR}(\alpha_{RI} \cdot r(T) + (1 - \alpha_{RI}) \cdot r(E) - r(E)) \end{aligned} \tag{11}$$

This leads us to predict that, under the same distribution of intermediary's donor-guilt types θ_{ID} and recipient-guilt types θ_{IR}, intermediaries' behavior should be the same in both treatments (between-subjects), as stated by our last behavioral hypothesis.

H.I4 [Donor *vs.* Recipient treatment on Choice]: Under the same distribution of psychological types, intermediaries' behavior is the same in both the Donor and the Recipient treatments.

3 Experimental Design and Procedures

We now describe in details how the game has been implemented in the laboratory.

3.1 Experimental Design

First-Order Belief Elicitation

In the first part of the experiment, we elicited the players' first-order beliefs about the donors' and the intermediaries' decisions in the game. Intermediaries and recipients had to report their beliefs about the number of donors, out of three donors randomly selected in the session, who choose to donate in the Low and in the High conditions that were played within-subjects. Similarly, donors and recipients had to report their beliefs about the number of intermediaries, out of three intermediaries randomly selected in the session, who choose to transfer the donation in full in each condition (conditional on the donor's decision to donate). The belief elicitation was incentivized. For each role, one belief was randomly selected at the end of the session and paid €1 if accurate.[16]

Donors' and Intermediaries' Decision-Making

In the second part of the experiment, subjects played the Embezzlement Mini-Game. Two treatments of this game were implemented between-subjects: the Donor treatment and the Recipient treatment. [17] Within-subjects, donors made a binary choice between giving a pre-determined fraction of their endowment and keeping their whole endowment, both in the Low and in the High conditions. These two decisions allow us to test whether the giving decision varies with the percentage potentially embezzled by the intermediary as predicted in Hypothesis H.D2.

[16]This incentivization procedure is the easiest to understand for subjects. Nevertheless, we contend that it is not perfectly incentive-compatible for risk-averse recipients who may under-estimate the probability that donors donate and that intermediaries transfer to the recipients. However, this concern is hindered both in theory – since they are four possible beliefs, one cannot be perfectly insured against risk – and in practice – we find an insignificant correlation between risk aversion and beliefs (see Table B.6 in Appendix).

[17]We used a between-subject design for studying the intermediaries' donor-guilt aversion and recipient-guilt aversion because we were anxious that using a within-subject design would be confusing for the subjects and could have created anchoring effects.

Meanwhile, intermediaries made binary choices between transferring the entirety of the amount given by the donor or transferring only a pre-determined fraction of this donation, both in the Low and in the High conditions. Whether intermediaries started with the Low or with the High condition was determined randomly at the individual level. These decisions were made under the veil of ignorance, *i.e.*, assuming that the donor had chosen to give a positive amount. We used the belief-dependent menu method of Khalmetski *et al.* (2015). In each condition, in the Donor (Recipient) treatment, intermediaries made four transfer decisions corresponding to the four possible first-order beliefs of the donor (recipient) on the frequency of intermediaries transferring: the donor's (recipient's) beliefs that none, one, two or three out of three intermediaries transfer in full. For facilitating decision-making, these first-order beliefs were presented in a fixed increasing order (see an example of a decision screen in Appendix D.1).[18]

At the end of the session, the computer program randomly selected either the Low or the High condition. Given the donor had given part of his endowment in this condition, the program implemented the intermediary's decision corresponding to the actual belief of the donor or of the recipient, depending on the treatment, in this condition. This determined the donor's, the intermediary's and the recipient's payoffs in this part.

Second-Order Belief Elicitation and Social Norms

[18]The use of the menu method is frequent in the experimental literature on guilt aversion (Attanasi *et al.*, 2013; Khalmetski *et al.*, 2015; Balafoutas and Fornwagner, 2017; Hauge, 2016; Bellemare *et al.*, 2017; Bellemare *et al.*, 2018, Dhami *et al.*, 2018). Although one might argue that responses elicited with this method are "cold", this method offers several advantages. First, it allows us to rule out potential false-consensus effects without raising the issue of strategic reporting and without using deception. The false-consensus effect could be avoided by communicating the donors' (recipients') true beliefs to the intermediaries. However, it requires choosing between two evils: if the donors (recipients) know that their beliefs will be communicated, they are likely to distort them; and if they do not know that their beliefs will be communicated, the design is arguably deceptive. The menu method avoids these drawbacks. Moreover, it allows us to study guilt aversion at the individual level and, hence, to unveil inter-individual differences that are hidden at the aggregate level (Khalmetski *et al.*, 2015).

In the third part of the experiment, we elicited the second-order beliefs of the donors and of the intermediaries on the other players' first-order beliefs, both in the Low and in the High conditions. Donors had to guess their intermediary's and their recipient's first-order beliefs on the donors' decisions (four second-order-beliefs in total). Similarly, intermediaries had to guess their donor's and their recipient's first-order beliefs on the intermediaries' decisions (four second-order-beliefs in total). A second-order-beliefs is considered correct if it corresponds to the partner's actual first-order belief.

Moreover, anticipating that behavior in this game may depend on social norms and on the beliefs about others' social norms, at the beginning of the third part we elicited all the subjects' social norms in the session and, at the end of this part, we elicited the donors' and the intermediaries' beliefs about their partners' social norms.[19]

The players' social norms were identified, using the Krupka and Weber's (2013) procedure, for each possible decision both in the Low and in the High conditions. In each condition, players had to rate the social appropriateness of each decision on a four-item scale (eight answers in total). An answer is considered correct if it corresponds to the modal answer of the subjects in the same role. Using coordination games among players with the same role to incentivize this procedure allows us to identify whether social norms differ across roles. In fact, similarly to Erkut *et al.* (2005), we found that social norms do not differ across roles in seven out of eight cases (Kruskal-Wallis tests, see Table B.7 in Appendix).[20]

Then, at the end of the third part, donors had to guess their intermediary's and their recipient's ratings of the social appropriateness of the donors' possible decisions (four

[19]Note that d'Adda *et al.* (2016) found no difference in responses between eliciting normative judgments à la Krupka and Weber before or after playing the main game.

[20]Ratings of social appropriateness differ in one instance only: in the Low condition, intermediaries consider that embezzling is less socially appropriate than donors do.

answers). Similarly, intermediaries had to guess their donor's and their recipient's ratings of the social appropriateness of the intermediaries' possible decisions (four answers). Recipients had no guess to report.

For each subject, we randomly selected one answer among all those provided during this third part. A correct answer paid €1.

Elicitation of Individual Characteristics

Since our model predicts that guilt proneness affects behavior in the game (Hypothesis HI3), we elicited the subjects' social preferences by means of several tests. A survey was completed online about a week prior the laboratory session to limit the risk of contamination between this task and the game. Subjects were paid a flat fee of €7 for completing this survey on time and for showing-up at the session in the laboratory.

The survey was composed of four parts (see Appendix D.2). In the first part, subjects completed the Guilt and Shame Proneness (GASP) questionnaire of Cohen *et al.* (2011). We were particularly attentive to the Guilt-Negative-Behavior-Evaluation subscale that assesses one's proneness to feel bad about how one acted. The second and third parts were included to control for potentially relevant psychological traits and to conceal our interest in guilt proneness. The second part corresponds to the Honesty-Humility scale extracted from the 100-item HEXACO Personality Inventory – Revised test (Ashton and Lee, 2008). We were interested in the responses to the Fairness subscale that aims at assessing a tendency to avoid dishonesty. The third part consists of 16 questions from the Self-Reported Altruism Scale (Rushton *et al.*, 1981). Finally, in the fourth part, we collected standard socio-demographic characteristics, including gender, age, professional status, number of past participations in economic experiments, self-reported risk attitudes (using the procedure of Dohmen *et al.*, 2011), and self-reported time preferences (using

the procedure of Visher *et al.*, 2013).

3.2 Procedures

The experiment was conducted at GATE-Lab, Lyon, France. It was computerized using the software Z-Tree (Fischbacher, 2007). Subjects were recruited mainly from the undergraduate student population of local business, engineering, and medical schools by email, using the software Hroot (Bock *et al.*, 2014). 369 subjects participated in a total of 19 sessions. 52.72% are females and the average age is 21.85 years (S.D. = 4.54). Table B.1 in Appendix summarizes the characteristics of each session.

When subjects registered for the experiment, about a week before the date of the lab session, they were sent an invitation email to complete the online questionnaire. Completing the questionnaire took about 10 minutes. Participants were informed that they would receive their fixed payment of €7 for this task and for showing-up at the end of the laboratory session. Only those who completed the online questionnaire were allowed to participate in the session. In the lab session, at their arrival subjects were randomly assigned to a cubicle after drawing a tag in an opaque bag. The instructions (see Appendix D.1) were distributed for each part after completion of the previous part. Before the first part, subjects had to answer a comprehension questionnaire. In the first part, subjects reported their first-order beliefs and donors made their decisions. In the second part, intermediaries made their decisions. In the third part, we elicited the subjects' social norms and second-order beliefs.

Each session lasted about 75 minutes. The average earnings were €17.70 (S.D. = 6.19), including the €7 fee for completing the online questionnaire and for showing-up. Earnings were paid in private in a separate room.

4 Results

We begin this section by two comments on the subjects' social norms and beliefs. Summary statistics and significance tests are displayed in Table B.2 in Appendix. First, *Give* and *Transfer* choices, respectively, are rated by the subjects as significantly more socially appropriate than *Keep* and *Embezzle* choices, respectively, in both conditions (Wilcoxon signed rank tests, W hereafter, $p < 0.001$). Second, the donors' second-order beliefs (SOB, hereafter) are accurate guesses of the intermediaries' and recipients' first-order beliefs (FOB, hereafter) on the frequency of *Give* choices in both conditions (Mann-Whitney rank sum tests, MW hereafter, between SOB and FOB, smallest $p = 0.44$). However, intermediaries tend to overestimate the donor's and recipient's FOB on the frequency of *Transfer* choices (MW tests, $p < 0.05$ in three out four cases).

In the following, we consider first the donors' behavior (Section 4.1) and next, the intermediaries' behavior (Section 4.2). We proceed by testing the behavioral hypotheses according to the order they have been introduced in Section 2.3. For each behavioral hypothesis, we also check for treatment differences, under the label [Donor *vs.* Recipient treatments].

Finally, in this paper, except when specified otherwise, the tests are two-sided, an independent observation corresponds to a decision (since only one decision per participant is payoff relevant) and the results from the two treatments are pooled.

4.1 Donors' Behavior

Our theoretical model is consistent with three patterns of donors' behavior: (i) *Keep* in both conditions (this is the case for 46.34% of the donors), (ii) *Give* in both conditions (31.70% of the donors), or (iii) *Give* only in the Low condition (17.07% of the donors).

Only 4.87% of the donors choose *Give* only in the High condition, a behavior inconsistent with our model. Figure 2 displays, for each condition, the proportion of donors who choose either *Give* or *Keep*, depending on their FOB on the frequency of *Transfer* choices (see also Table B.3 in Appendix). It illustrates our two results on the donors' behavior.

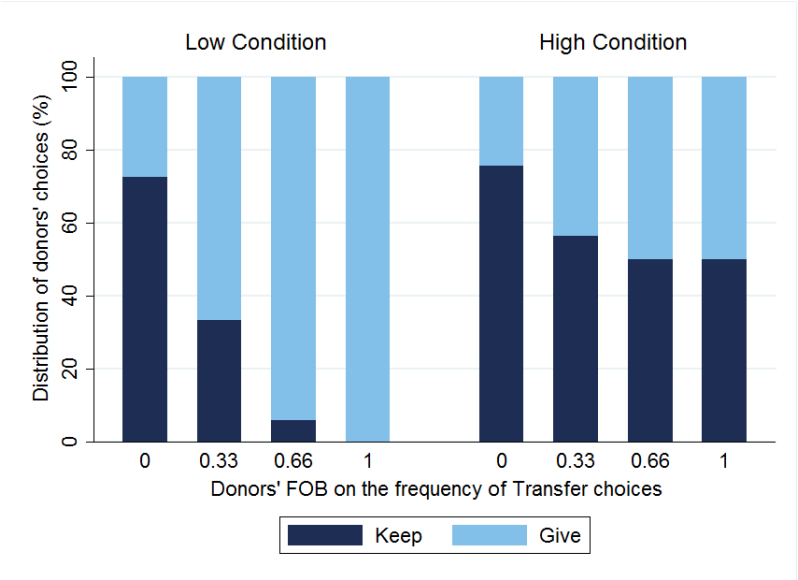

Figure 2: Distribution of the donors' choices depending on their first-order beliefs
Note: One donor had a FOB of 1 in the Low condition; as well as two donors in the High condition.

Result D1 [Choice-Belief Correlation]: The higher the donors' FOB about *Transfer*, the higher the frequency of *Give* choices. This holds for both conditions.

Support for Result D1: There is a significant positive correlation between the donors' FOB about *Transfer* and their decision to *Give* (Spearman rank correlation, S hereafter, $r_s = 0.35$, $p < 0.001$). When we distinguish between conditions, the correlation in the Low condition (S correlation, $r_s = 0.51$, $p < 0.001$) is significantly higher than in the High condition (S correlation, $r_s = 0.22$, $p < 0.001$) (ZPF statistic, $z = 3.05$, $p < 0.001$).[21]

[21]The correlation between the donors' FOB and their decision to *Give* must be regarded with caution.

[Donor *vs.* Recipient treatments]: The correlation between the donors' FOB on the frequency of *Transfer* choices and their decision to *Give* is not different across treatments (Recipient treatment: $r_s = 0.24$, $p < 0.001$; Donor treatment: $r_s = 0.44$, $p < 0.001$; Z test, $z = -1.23$, $p = 0.210$).

Result D2 [High *vs.* Low Condition on Choice]: Controlling for the donors' FOB about *Transfer*, the frequency of *Give* choices is higher in the Low than in the High condition.

Support for Result D2: We use Mc Nemar tests (MN, hereafter) to consider each donor as an independent observation. For a given FOB about *Transfer*, the frequency of *Give* choices is significantly higher in the Low than in the High condition (MN tests; FOB(0): χ^2=4.76, $p = 0.029$; FOB (0.33): χ^2=3.60, $p = 0.057$).[22]

[Donor *vs.* Recipient treatments]: Even though donors could not know which treatment is implemented when they made their choices, our results differ across treatments. Result D2 is supported in the Donor treatment (MN tests; FOB(0): χ^2=3.57, $p = 0.058$; FOB (0.33): χ^2=3.00, $p = 0.083$) but not in the Recipient Treatment (MN tests; FOB(0): χ^2=1.60, $p = 0.205$; FOB (0.33): χ^2=1.29, $p = 0.252$).

Although belief elicitation was incentivized, it is possible that donors who planned to *Keep* may have underestimated their FOB about *Transfer* to justify their selfish choice. We consider the donors' rating of the social appropriateness of embezzling as a proxy for their FOB on the frequency of *Transfer* choices because (i) they are significantly correlated (S correlation, $r_s = -0.19$, $p < 0.001$), and (ii) we believe that it is more unlikely that donors used their rating of the social appropriateness of embezzling, rather then their FOB, as a justification of their choice. We replicate the correlation with the ratings of the social appropriateness of embezzling (S correlation, $r_s = -0.20$, $p < 0.001$).

[22]No donor had a FOB of either 0.66 in both conditions or 1 in both conditions.

4.2 Intermediaries' Behavior

Our theoretical model is consistent with two patterns of intermediaries' behavior: (i) always choose *Embezzle*, which characterizes selfish intermediaries, and (ii) switching from choosing *Embezzle* to choosing *Transfer* as the induced SOB increases, which characterizes guilt-averse intermediaries. Our model captures more than 70% of the observed behavior: on average, 46.75% of the intermediaries are selfish and 24.39% are guilt-averse when we pool the two treatments. The remaining intermediaries behave as follows: 11.79% always choose *Transfer*, thereby exhibiting belief-independent social preferences; 11.79% switch multiple times between transferring and embezzling; 5.28% exhibit an inverse switching pattern from transferring to embezzling.

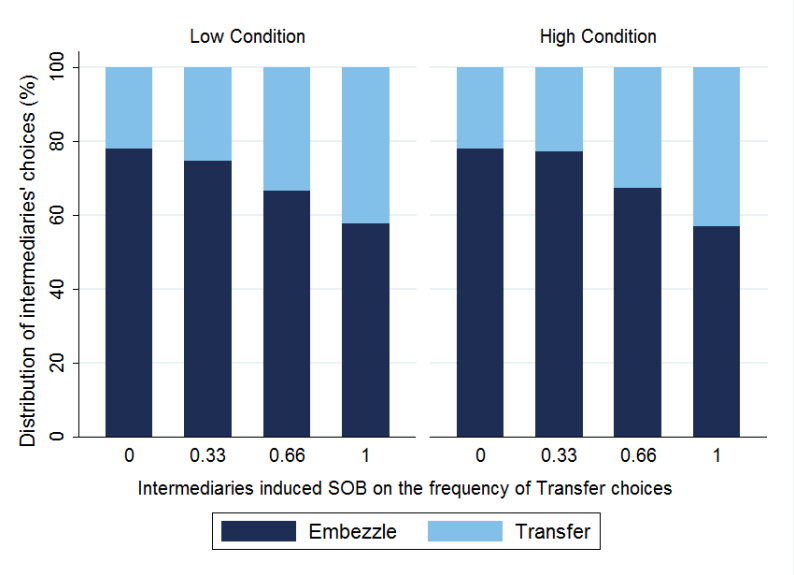

Figure 3: Distribution of the intermediaries' choices depending on their induced second-order beliefs

Figure 3 displays, for each condition, the proportion of intermediaries who choose either *Transfer* or *Embezzle*, depending on their induced SOB (see also Table B.4 in Appendix). It illustrates our two results on the intermediaries' behavior.

Result I1 [Choice-Belief Correlation]: The higher the intermediaries' SOB about *Transfer*, the higher the frequency of *Transfer* choices. This holds independently of the condition.

Support for Result I1: There is a significant positive correlation between the intermediaries' induced SOB about *Transfer* and their *Transfer* choice (S correlation, r_s=0.15, p <0.001). The correlation does not vary between conditions (Low: r_s= 0.17, p <0.001; High: $r_s = 0.18$, p <0.001). Note that, if we exclude the intermediaries who believed that no donor would *Give* in either condition, the correlation increases to $r_s = 0.22$ (p <0.001). Indeed, these excluded intermediaries may suffer from an hypothetical bias, as they believe that their choices will not be payoff-relevant, rendering the hypothetical decision to embezzle less psychologically costly. Interestingly, if we correlate the intermediaries' *Transfer* choices with their stated SOB rather than with their induced SOB, the correlation increases to $r_s = 0.27$ (p <0.001).[23] We also find support for Result I1 using a Logit model with fixed effects (Table 1) and with random effects as well as individuals controls (Table B.5 in Appendix).

[Donor *vs.* Recipient treatments]: The correlation between the intermediaries' induced SOB about *Transfer* and their *Transfer* choices does not vary significantly across treatments (Donor treatment: r_s=0.14, p <0.001; Recipient treatment: $r_s = 0.15$, p <0.001; Z-test, z=0.05, p=0.95) (see also Table 1 in Appendix).

Result I2 [High *vs.* Low condition on Choice]: Controlling for the intermediaries' SOB about *Transfer*, the frequency of a *Transfer* choices does not differ across conditions.

[23]We interpret this increase as evidence of a false-consensus effect (Ross, 1977; Vanberg, 2008). Experiments using stated SOB should not ignore this effect as it leads to an upward-bias measure of the correlation between SOB and choices (see consistent results in Khalmetski *et al.*, 2015; and Bellemare *et al.*, 2017a).

	All treatments	Recipient treatment	Donor treatment	Hypothetical Bias Excluded	All treatments
Induced SOB	0.67***	0.75***	0.62***	0.81***	
	(0.09)	(0.14)	(0.12)	(0.11)	
Low Condition	0.06	-0.13	0.21	0.10	-0.24
	(0.20)	(0.30)	(0.26)	(0.22)	(0.52)
Stated SOB					1.27**
					(0.63)
# Observations	472	216	256	400	42
# Participants	59	27	32	50	21

Notes: In the four first columns, the dependent variable is the decision to *Transfer* made for a given induced SOB. In the last column, the dependent variable is the decision to *Transfer* made when the induced SOB corresponded to the stated SOB. Standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01

Table 1: Regression on the decision to Transfer (Logit model with fixed effects)

Support for Result I2: We use MN tests to consider each intermediary as an independent observation. For a given induced SOB, the frequency of a *Transfer* choices does not significantly differ across conditions (smallest $p = 0.438$) (see also Table 1 in Appendix).

[Donor *vs.* Recipient treatments]: We replicate this result when we distinguish between the Donor and the Recipient treatments in seven out of eight cases (MN tests for each induced SOB, smallest $p = 0.256$), with one exception (Recipient treatment when SOB=0.33: χ^2=4.50, $p = 0.033$) (see also Table 1 in Appendix).

Result I3 [Choice-Guilt Correlation]: The higher the guilt-sensitivity, (i) the higher the frequency of a *Transfer* choice controlling for intermediaries' SOB; (ii) the lower the minimum SOB sufficient to switch from *Embezzle* to *Transfer*. This holds regardless of the condition.

Support for Result I3: We consider the Guilt-Negative-Behavior-Evaluation score (Guilt-NBE score, hereafter) elicited in the pre-experimental survey as a proxy for the

guilt-sensitivity parameter in our model. Table 2 presents (i) the correlation between the *Transfer* choices, holding the induced SOB constant, and the Guilt-NBE score, as well as (ii) the correlation between the switching SOB and the Guilt-NBE score. The switching SOB corresponds to the minimum induced SOB sufficient to choose *Transfer* rather than *Embezzle*.[24] The Guilt-NBE score in itself is only marginally significantly correlated with *Transfer* choices. However, the significance of this correlation can be improved if we interact this score with the Fairness score (which measures a tendency to avoid dishonesty) or with a dummy variable that takes value 0 if the intermediary believes that no donor will choose Give in either condition, and 1 otherwise. This suggests that the Guilt-NBE score is a relevant proxy for the guilt-sensitivity parameter but only when intermediaries already have a high moral stand on corruption or when they are sure their decision will be implemented.

[Donor *vs.* Recipient treatments]: The magnitude of this correlation is lower in the Donor treatment than in the Recipient treatment, but not significantly so (Z-tests, smallest $p = 0.14$).

Result I4 [Donor *vs.* Recipient]: All our hypotheses hold independently of whether guilt is directed toward the donor or toward the recipient.

Support for Result I4: For each result, see (the absence of) treatment difference under the label [Donor *vs.* Recipient treatments].

[24] For an intermediary who always Transfers, the swithching SOB is 0 ; for an intermediary who Embezzles when the induced SOB is 0 and Transfers when the induced SOB is $\{0.33; 0.66; 1\}$, the switching SOB is 0.33 ; etc. We cannot compute a switching SOB for intermediaries who exhibited multiples switches or an inverse switching pattern.

	Recipient treatment			Donor treatment		
	Guilt	Guilt x Fairness	Guilt x Hypoth.	Guilt	Guilt x Fairness	Guilt x Hypoth.
Transfer | SOB=0	0.17	0.27**	0.23*	0.10	0.16	0.27**
Transfer | SOB=0.33	0.43***	0.49***	0.46***	0.19	0.28**	0.39***
Transfer | SOB=0.66	0.19*	0.26**	0.43***	0.15	0.22*	0.47***
Transfer | SOB=1	0.23*	0.31**	0.52***	0.11	0.14	0.49***
Switching SOB	-0.10	-0.20*	-0.34***	-0.14	-0.16	-0.49***

Notes: This table presents the Spearman correlation between row and column variables.
Standard errors in parentheses; * p <0.1, ** p <0.05, *** p <0.01.
Rows: The first four lines represent the total of *Transfer* choices given each induced FOB: 0 if *Transfer* in neither condition, 1 if *Transfer* in one condition, 2 if *Transfer* in both conditions. The fifth line represents the switching SOB of intermediaries who either always embezzle or are guilt-averse.
SOB stands for second-order beliefs. Guilt stands for Guilt-NBE score (GASP questionnaire). Fairness stands for fairness score (HEXACO questionnaire). Hypoth. stands for a dummy variable that takes value 0 if the intermediary believes no donor will choose *Give* in either condition, and 1 otherwise.

Table 2: Correlation between the intermediaries' decisions and their Guilt-NBE score

4.3 A Structural Estimate of Guilt Sensitivity

Figure 4 presents the distribution of the switching SOB observed in the two treatments.

[Donor *vs.* Recipient treatments]: The distributions of switching SOB do not differ significantly across treatments (Kruskal-Wallis test, p >0.10).

Following Bellemare *et al.* (2011), we defined a structural econometric model (Eq. 12) to estimate the intermediaries' average guilt-sensitivity parameter, θ_{Ij}, toward the donor ($j = D$, in the Donor treatment) and the receiver ($j = R$, in the Donor treatment). Equation 12 folows our modeling of the intermediary's guilt aversion toward the donor and the recipient (Eq. 5). The utility of an intermediary I who takes a decision $s_I \in \{Transfer, Embezzle\}$ is composed of two terms. First, it depends on the intermediary's material payoff when choosing $s_I : M_I(s_I)$. Second, it depends on the intermediary's guilt toward the other player j when choosing $s_I : G_{Ij}(s_I)$, with $j = D, R$. The intermediary's

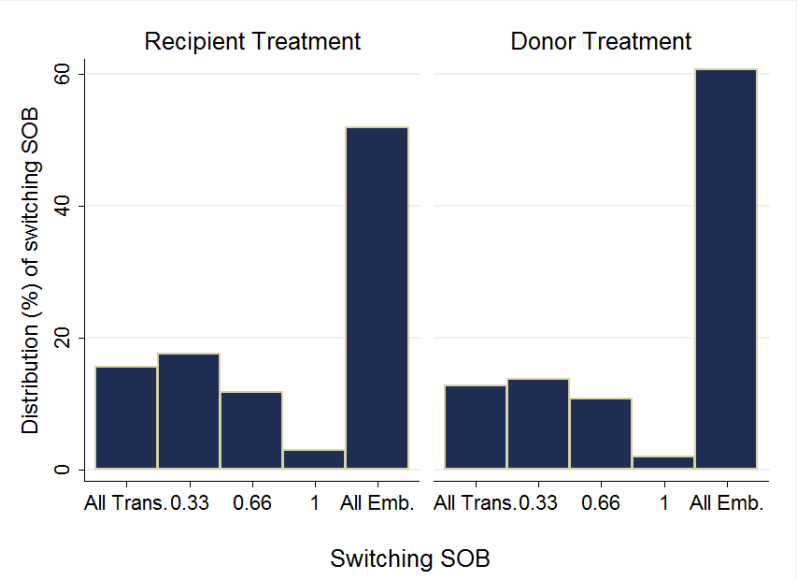

Figure 4: Distribution of the intermediaries' switching second-order beliefs
Note: For instance, in the Recipient Treatment, 3% of the intermediaries have a switching SOB of 1 (i.e. chose *Embezzle* for induced SOB of $\{0; 0.33; 0.66\}$ and *Transfer* for an induced SOB of 1) and 51% of intermediaries always Embezzled irrespective of the induced SOB.

guilt is the difference, if positive, between the other player's (j) first-order belief (α_{jI}) on the recipient's material payoff and the recipient's actual payoff.

$$\begin{aligned} U_I(s_I, \alpha_{jI}, \theta_{Ij}) &= M_I(s_I) + \theta_{Ij} G_{Ij}(s_I) \\ &= M_I(s_I) + \theta_{Ij}[\alpha_{jI}(T) \cdot M_R(T) + (1 - \alpha_{jI}(T)) \cdot M_R(E) - M_R(s_I)] \quad (12) \end{aligned}$$

We used a conditional Logit model to estimate θ_I, the coefficient corresponding to $G_{Ij}(s_I)$, while fixing to 1 the coefficient corresponding to $M_I(s_I)$. The results reported in Table 3 show that the average intermediary is willing to give up embezzling 0.12 ECU if embezzling increases another player's disappointment (difference between expectations and actual outcome) by 1 ECU. When we exclude intermediaries who believed that no donor chose to *Give* (those potentially subject to an hypothetical bias), the estimated guilt-sensitivity parameter increases up to 0.20.

[Donor *vs.* Recipient treatments]: Comparing the estimated parameter across treatments, we find a significant difference. Intermediaries seems to be slightly more sensitive to recipient's guilt (+18%) (Z-test[25], χ^2=-2, p=0.04).

	All treatments	Recipient treatment	Donor treatment	Subjects with Hypothetical Bias Excluded
θ_I	-0.12***	-0.13***	-0.11***	-0.20***
	(0.01)	(0.01)	(0.01)	(0.01)
N	123	62	61	83

Notes: Standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01

Table 3: Structural estimates of the guilt-sensitivity parameter

5 Discussion and Conclusion

In this study we investigated theoretically and experimentally the role of guilt aversion in the behavior of intermediaries confronted with an opportunity to embezzle a donation. Using psychological game theory, our aim was to determine (i) whether others' expectations influence the decision to embezzle, and if so (ii) whether the impact of others' expectations on behavior is different if others are the donors or the potential beneficiaries of the donation. Extending Battigalli and Dufwenberg (2007) model to capture guilt aversion toward the donor and documenting its existence and features by means of a laboratory experiment are the two original contributions of our paper. Indeed, beyond showing evidence of guilt-averse behavior, the recent experimental literature on guilt aversion has often pursued three separate objectives: measuring the prevalence of guilt aversion in the population and its magnitude, and identifying a survey-based measure of guilt aversion. To our knowledge, we are the first to address these three questions in a

[25]See Paternoster *et al.* (1998).

single paper. More importantly, we also consider a new direction of guilt whose existence was not documented yet: guilt directed toward a player whose payoffs cannot be affected by the agent's decision.

We found that (i) on average, about 25 % of intermediaries are affected by others' expectations in the way predicted by our guilt-aversion model, and the proportion of guilt-averse intermediaries is not affected by the direction of the guilt; (ii) on average, an intermediary is willing to give up embezzling 0.12 ECU if embezzling increases another player's disappointment by 1 ECU, but the intensity of the structurally estimated guilt-sensitivity parameter is higher when the intermediary is confronted with the recipient's expectations (0.13) compared to the donor's ones (0.11).

Thus, our results extend the recent strand of the literature aiming at estimating the proportion of guilt-averse individuals in the population that was so far limited to Dictator games (see Table C.1 in Appendix). Our structural estimates of guilt sensitivity are in the lower range of values previously obtained through structural or equilibrium estimations (see Table C.2 in Appendix). A possible interpretation is that since our game involves more than one active role, the intermediaries' feeling of responsibility is more diluted than in games in which only one player is active. Finally, we report a significant positive correlation between the intermediaries' switching second-order beliefs and their Guilt-Negative-Behavior-Evaluation score, but only when intermediaries already have high moral standards. This finding contributes to the small but inclusive literature trying to identify the link between survey-based measures of guilt and experimental decisions (see Table C.3 in Appendix). Overall, these contrasted results call for more research on the nature of the emotions embedded in the Battigalli and Dufwenberg (2007) guilt-aversion model.

Showing evidence that intermediaries care about others' expectations suggests that

anti-corruption policies could benefit from developing interventions making donors' and recipients' expectations of good behavior more salient to the intermediaries, promoting especially the high expectations others hold. We already know that public campaigns of information (Reinikka and Svensson, 2011) or framing manipulations (Ockenfels and Werner, 2014) affect decisions through beliefs. But so far, attention has been focused on the potential recipients' expectations. Our findings show that future policies should also consider the sensitiveness of intermediaries to the donors' expectations (*e.g.*, tax payers, charitable donors) (see also the literature on trust-responsiveness, *e.g.*, Guerra and Zizzo, 2004; Bacharrach *et al.*, 2007). Another implication of our findings is related to social norms. By enlarging the perspective to a dynamic setting, as in Balafoutas (2011), we could contribute to explain the "vicious circle" of corruption. If donors or recipients expect a high level of embezzlement in the society, intermediaries may embezzle without feeling guilty, which in turn increases the expectations of embezzlement.

We measured guilt aversion toward the donor and toward the recipient in two separate treatments. A straightforward extension would be to test a treatment in which intermediaries would be informed about both donors' and recipients' expectations. This would lead to a more complex design, though. Finally, one may worry about the external validity of our findings since our experiment has been conducted in an artificial setting in a developed country. We remain confident that our experiment captures some important features of the real-life environment, such as the significant difference between the social appropriateness of transferring *vs.* embezzling, or the fact that donors condition their donation on their belief that intermediaries will transfer it honestly (see also Chlaß *et al.*, 2015). Moreover, in a bribery experiment conducted both in the field in the lab, in Burkina Faso and in Canada, Armantier and Boly (2013) reported no significant difference in behavior across contexts. Still, future research could test whether our results replicate in

the field. A major challenge, though, will be to measure beliefs in the field. We leave this to future investigations.

References

Abbink, K., Serra, D. (2012). Anticorruption Policies: Lessons from the Lab. *New Advances in Experimental Research on Corruption 15*, 77–115.

Abeler, J., Nosenzo, D., Raymond, C. (2018). Preferences for truth-telling. *Econometrica*. Forthcoming.

D'Adda, G., Drouvelis, M., Nosenzo, D. (2016). Norm elicitation in within-subject designs: Testing for order effects. *Journal of Behavioral and Experimental Economics 62*, 1–7.

Armantier, O., Boly, A. (2013). Comparing Corruption in the Laboratory and in the Field in Burkina Faso and in Canada. *The Economic Journal 123(573)*, 1168-1187.

Ashton, M.C., Lee, K. (2008). The prediction of Honesty-Humility-related criteria by the HEXACO and Five-Factor Models of personality. *Journal of Research in Personality 42(5)*, 1216–1228.

Attanasi, G., Battigalli, P., Manzoni, E. (2016). Incomplete Information Models of Guilt Aversion in the Trust Game. *Management Science 62(3)*, 648–667.

Attanasi, G., Battigalli, P., Nagel, R. (2013). Disclosure of Belief-dependent Preferences in a Trust Game. *IGIER Working paper No. 506*.

Attanasi, G., Battigalli, P., Manzoni, E., Nagel, R. (2018). Belief-dependent Preferences and Reputation: Experimental Analysis of a Repeated Trust Game. *IGIER Working paper No. 622*.

Bacharach, M., Guerra, G., Zizzo, D. J. (2007). The self-fulfilling property of trust: An experimental study. *Theory and Decision 63(4)*, 349–388.

Bahr, G., Requate, T. (2013). Reciprocity and Giving in a Consecutive Three-Person Dictator Game with Social Interaction. *German Economic Review 15(3)*, 374–392.

Balafoutas, L. (2011) : Public beliefs and corruption in a repeated psychological game. *Working Papers in Economics and Statistics*, No. 2009-01.

Balafoutas, L., Fornwagner, H. (2017). The Limits of Guilt. *Journal of the Economic Science Association 3(2)*, 137–148.

Barr, A., Lindelow, M., Serneels, P. (2009). Corruption in public service delivery: An experimental analysis. *Journal of Economic Behavior and Organization 72(1)*, 225–239.

Battigalli, P., Siniscalchi, M. (2002). Strong belief and forward induction reasoning, *Journal of Economic Theory*, 106, 356-391.

Battigalli, P., Dufwenberg, M. (2007). Guilt in Games. *American Economic Review 97(2)*, 170–176.

Battigalli, P., Dufwenberg, M. (2009). Dynamic psychological games. *Journal of Economic Theory 144(1)*, 1–35.

Battigalli, P., Corrao, R., Sanna, F. (2017). Epistemic game theory without type structures. An application to psychological games. Typescript, Bocconi University.

Becker, G.S., Stigler, G.J. (1974). Law enforcement, malfeasance, and the compensation of enforcers. *Journal of Legal Studies 3*, 1–18.

Bellemare, C., Sebald, A., Strobel, M. (2011). Measuring the willingness to pay to avoid guilt: estimation using equilibrium and stated belief models. *Journal of Applied Econometrics*, 26(3), 437-453.

Bellemare, C., Sebald, A., Suetens, S. (2017). A note on testing guilt aversion. *Games and Economic Behavior 102*, 233–239.

Bellemare, C., Sebald, A., Suetens, S. (2018). Heterogeneous guilt sensitivities and incentive effects. *Experimental Economics*, Forthcoming.

Bock, O., Baetge, I., Nicklisch, A. (2014). Hroot: Hamburg Registration and Organization Online Tool. *European Economic Review 71*, 117-120.

Bolton, G. E., Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review 90(1)*, 166-193.

Boly, A., Gillanders, R., Miettinen, T. (2016). Deterrence, peer effect and legitimacy in anti- corruption policy-making. An experimental analysis. *WIDER Working Paper No. 137.*

Bracht, J., Regner, T. (2013). Moral emotions and partnership. *Journal of Economic Psychology 39*, 313-326.

Canagarajah, S., Ye, X. (2001). Public Health and Education Spending in Ghana in 1992-98: Issues of Equity and Efficiency. *World Bank Publications*.

Charness, G., Dufwenberg, M. (2006). Promises and Partnership. *Econometrica 74*, 1579–1601.

Charness, G., Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics, 117(3)*, 817-869.

Chlaß, N., Gangadharan, L., Jones, K. (2015). Charitable Giving and Intermediation. *Jena Economics Research Papers*.

Cohen, T.R., Wolf, S.T., Panter, A.T., Insko, C.A. (2011). Introducing the GASP scale: a new measure of guilt and shame proneness. *Journal of Personality and Social Psychology 100(5)*, 947–966.

Dahmi, S., Wei, M. al-Nowaihi, A. (2018). Public goods games and psychological utility: Theory and evidence. *Forthcoming in Journal of Economic Behavior and Organization*.

Di Falco, S., Magdalou, B., Masclet, D., Villeval, M.C., Willinger, M. (2016). Can transparency of information reduce embezzlement? Experimental Evidence from Tanzania. *SSRN Working Paper*, http://ssrn.com/abstract=2775659.

Di Tella, R., Schargrodsky, E. (2003). The Role of Wages and Auditing during a Crackdown on Corruption in the City of Buenos Aires. *Journal of Law and Economics 46(1)*, 269–292.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G. (2011). Individual Risk Attitudes: Measurements, Determinants and Behavioral Consequences. *Journal of the European Economic Association 9(3)*, 522–550.

Drugov, M., Hamman, J., Serra, D. (2014). Intermediaries in corruption: An experiment. *Experimental Economics 17(1)*, 78–99.

Erkut, H., Nosenzo, D., Sefton, M. (2015). Identifying social norms using coordination games: Spectators vs. stakeholders. *Economics Letters 130*, 28-31.

Fan, C.S., Lin, C., Treisman, D. (2009). Political decentralization and corruption: Evidence from around the world. *Journal of Public Economics 93(1)*, 14–34.

Fehr, E., Schmidt, K. (1999). A Theory of Faimess, Competition, and Cooperation. *The*

Quarterly Journal of Economics 114, 817-868.

Ferraz, C., Finan, F., Moreira, D.B. (2012). Corrupting learning: Evidence from missing federal education funds in Brazil. *Journal of Public Economics 96(9)*, 712–726.

Fischbacher, U. (2007). Ztree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics 10(2)*, 171–178.

Fischbacher, U., Föllmi-Heusi, F. (2013). Lies in disguise: An experimental study on cheating. *Journal of the European Economic Association 11(3)*, 525–547.

Gaitonde, R., Ad, O., Po, O., Rada, G. (2016). Interventions to reduce corruption in the health sector (Review). *Cochrane Database of Systematic Reviews (8)*.

Geanakoplos, J., Pearce, D., Stacchetti, E. (1989). Psychological Games and Sequential Rationality. *Games and Economic Behavior 1(1)*, 60–79.

Guerra, G., Zizzo, D. J. (2004). Trust responsiveness and beliefs. *Journal of Economic Behavior and Organization 55(1)*, 25-30.

Hauge, K.E. (2016). Generosity and guilt: The role of beliefs and moral standards of others. *Journal of Economic Psychology* 54, 35–43.

Khalmetski, K., Ockenfels, A., Werner, P. (2015). Surprising gifts: Theory and laboratory evidence. *Journal of Economic Theory 159*, 163–208.

Krupka, E. L., Weber, R. A. (2013). Identifying social norms using coordination games: why does dictator game sharing vary? *Journal of the European Economic Association 11*, 495–524.

Mazar, N., Amir, O., Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research 45(6)*, 633-644.

Moulton, R. W., Burnstein, E., Liberty Jr, P. G., Altucher, N. (1966). Patterning of parental affection and disciplinary dominance as a determinant of guilt and sex typing. *Journal of Personality and Social Psychology 4(4)*, 356.

Ockenfels, A., Werner, P. (2014). Scale manipulation in dictator games. *Journal of Economic Behavior and Organization 97*, 138-142.

Olken, B. A. (2007). Monitoring Corruption: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy 115(2)*, 200–249.

Olken, B.A., Pande, R. (2012). Accessed Corruption in Developing Countries. *MIT Annual Reviews*.

Patel, A., Smith, A. (2018). Guilt and participation. *School of Economics Working Paper, University of East Anglia* (No. 2018-01).

Paternoster, R., Brame, R., Mazerolle, P., Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology 36(4)*, 859-86

Peeters, R., Vorsatz, M. (2018). Simple guilt and cooperation. *University of Otago Economics Discussion Papers* (No. 1801).

Regner, T., Harth, N. S. (2014). Testing belief-dependent models. *Max-Planck Institute of Economics Jena Working Paper.*

Reinikka, R., Svensson, J. (2004). Evidence from a Central Government Transfer Program in Uganda. *The Quarterly Journal of Economics 119(2)*, 679–705.

Reinikka, R., Svensson, J. (2011). The power of information in public services: Evidence from education in Uganda. *Journal of Public Economics 95(7-8)*, 956-966.

Ross, L. (1977). The false consensus effect: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13(3)*, 279-301.

Rushton, J. P., Chrisjohn, D. R., Fekken, G. C. (1981). The Altruistic Personality and the Self-Report Altruism Scale. *Personality and Individual Differences 2(4)*, 293–302.

Tangney, J.P., Fischer, K.W. (1995). *Self-conscious emotions: The psychology of shame, guilt, embarrassment and pride.* New York: Guilford P.

Vanberg, C. (2008). Why do people keep their promises? An experimental test of two explanations. *Econometrica 76(6)*, 1467-1480.

Vischer, T., Dohmen, T., Falk, A., Huffman, D., Schupp, J., Wagner, G.G. (2013). Validating an ultra-short survey measure of patience. *Economics Letters 120(2)*, 142–145.

Appendices

A Psychological Type Structure

The psychological type of the donor is characterized by his sensitivity to altruism, while the one of the intermediary is characterized by his sensitivity to guilt. Therefore, we can build a Harsanyi-type structure (as in Attanasi *et al.*, 2016) where each subject in role $j = D, I, R$ is identified by his **type** t_j, which comprises his psychological type and his exogenous beliefs about the type of the co-player (exogenous higher-order beliefs). Without loss of generality, the set of types may be assumed to be a Cartesian product $T_j = p_j \times e_j$, so that a type is a pair (p_j, e_j) of *p*sychological and *e*pistemic type, and the exogenous beliefs about the co-players' types are determined only by e_j, the epistemic type. We assume that the psychological and epistemic types of each player are statistically independent, and that this is common knowledge. We make the auxiliary assumption that in the Donor (Recipient) treatment the recipient's (donor's) type is not considered in the type structure (see discussion at the end of Section 2.2.4). With this, in the Donor treatment, a donor's type is $t_D = (\gamma_D, e_D)$ and an intermediary's type is $t_I = (\theta_{ID}, e_{ID})$, where e_D parametrizes the subjective probability assigned by the donor to $\bar{t}_{ID} = (\bar{\theta}_{ID}, e_{ID})$ – the positive-guilt type of the intermediary –, and e_{ID} parametrizes the subjective probability assigned by the intermediary to $\bar{t}_D = (\bar{\gamma}_D, e_D)$ – the altruistic type of the donor. In the Recipient treatment, because the recipient has only one possible psychological type (he is selfish), the Harsanyi and the epistemic types coincide, $t_R = e_R$, where e_R parametrizes the subjective probability assigned by the recipient to $\bar{t}_{IR} = (\bar{\theta}_{IR}, e_{IR})$, and e_{IR} parametrizes the intermediary's second-order exogenous beliefs over $\bar{\theta}_{IR}$, via e_R.

Differently from Attanasi *et al.* (2016), we do not use Bayesian equilibrium as solution concept. We instead rely on **two-step rationalizability with forward induction**: rationality (first step) and strong belief in rationality (second step). As highlighted by Attanasi *et al.* (2013), this procedure provides coarse behavioral predictions with respect to a Bayesian-equilibrium approach. However, it has a main methodological advantage for psychological games with incomplete information: since subject j's utility of outcomes only depends on j's own personal traits (and possibly on the co-player's beliefs), the analysis of players' rationality is independent of whether there is complete or incomplete information. The same is true, under mild assumptions on the donor's psychological type, for the analysis of strong belief in rationality by the intermediary. This allows us to elaborate behavioral predictions without positing any specific assumption concerning players' epistemic types, *i.e.*, over the **exogenous beliefs** about the co-player's (first-order exogenous beliefs) and their own (second-order exogenous beliefs) psychological type. This is crucial especially in an experimental environment where exogenous beliefs are not elicited (where, *e.g.*, donors are not asked to guess how many intermediaries are guilt-averse). Indeed, we only elicit and elaborate predictions on **endogenous beliefs**, *i.e.*, about players' behavior, which can be later observed during the experiment. To the best of our knowledge, not eliciting exogenous beliefs is common to all other experiments on games with belief-dependent preferences.

B Additional Tables

Session (#)	Participants (n)	Age (mean)	Previous Exp. (mean)	Economics Stud. (%)
		Donor Treatment		
5	18	21.27	1.44	66.67
6	18	21.00	0.22	33.33
7	15	24.2	1.53	33.33
8	21	22.00	0.80	28.57
9	21	21.28	1.80	71.43
12	24	23.25	1.33	45.83
13	18	20.66	1.50	61.11
15	21	21.19	1.00	38.10
16	12	21.00	2.50	50.00
18	15	22.4	1.86	33.33
Sub-total	183	21.83	1.34	46.45
		Recipient Treatment		
1	18	21.77	1.16	77.78
2	21	19.76	0.90	57.14
3	15	20.93	0.26	80.00
4	21	20.85	1.14	52.38
10	18	22.88	1.16	61.11
11	27	22.30	1.96	62.96
14	24	21.50	2.20	54.17
17	27	24.59	2.70	55.56
19	15	21.00	2.46	46.67
Sub-total	186	21.87	1.63	60.22
Treatment Difference		No[1]	No[1]	Yes[2]***
Total	369	21.85	1.49	53.39

Notes: Standard errors in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01
[1] Mann-Whitney ranks sum tests; [2] Fisher exact test

Table B.1: Summary statistics of participants per session

	Low	z-stat	High	z-stat
On the donors' behavior				
Intermediaries' FOB on the frequency of Give choices [a]	0.39	0.42	0.37	-0.21
Donors' SOB on intermediaries' FOB [a]	0.40		0.35	
Recipients' FOB on the frequency of Give choices [a]	0.40	-0.76	0.36	-0.30
Donors' SOB on recipients' FOB [a]	0.37		0.34	
Social Norm on Send [b]	0.88	-16.10***	0.84	-15.78***
Social Norm on Keep [b]	-0.48		-0.43	
On the intermediaries' behavior				
Donors' FOB on the frequency of Transfer choices [a]	0.20	-4.72***	0.25	-2.14**
Intermediaries' SOB on donors' FOB [a]	0.36		0.27	
Recipients' FOB on the frequency of Transfer choices [a]	0.21	3.15**	0.27	1.24
Intermediaries' SOB on recipients' FOB [a]	0.30		0.29	
Social Norm on Transfer [b]	0.89	-14.45***	0.90	-15.74***
Social Norm on Embezzle [b]	0.19		-0.18	

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

[a] Average beliefs on the frequency of choices are rated on scale from 0 (never) to 1 (always). Differences between FOB and SOB are measured by Mann-Whitney rank sum tests.

[b] Average social norms are rated on a scale from -1 (very socially inappropriate) to 1 (very socially appropriate). Differences between social norms are measured by Wilcoxon signed rank tests.

Table B.2: Summary statistics on beliefs and social norms

	Low Condition		High Condition	
	%	n	%	n
Give | FOB=0	27.54%	19	24.53%	13
Give | FOB=0.33	66.67%	24	43.75%	21
Give | FOB=0.66	94.12%	16	50.00%	10
Give | FOB=1	100%	1	50.00%	1

Notes: For each condition, a donor makes one choice given his FOB, e.g., in the Low condition, among the donors whose FOB was 0.33, 66.67% chose *Give*.

Table B.3: Donors' *Give* choices for a given FOB on the frequency of *Transfer* choices

	Low Condition		High Condition	
	%	n	%	n
Transfer | SOB=0	21.95%	27	21.95%	27
Transfer | SOB=0.33	25.20%	31	27.76%	28
Transfer | SOB=0.66	33.33%	41	32.52%	40
Transfer | SOB=1	42.28%	52	43.09%	53

Notes: For each condition, an intermediary makes four choices given each induced SOB, e.g., in the Low condition, when the induced SOB was 0.33, 25.20% of intermediaries chose *Transfer*.

Table B.4: Intermediaries' *Transfer* choices for a given induced SOB

	All treatments	Recipient treatment	Donor treatment	Hypothetical Bias Excluded	All treatments
Induced SOB	0.75***	0.76***	0.74***	0.90***	
	(0.10)	(0.15)	(0.13)	(0.11)	
Low Condition	0.08	-0.13	0.27	0.13	-0.04
	(0.20)	(0.30)	(0.28)	(0.23)	(0.47)
Donor Treatment	-0.94			-0.54	-0.93
	(0.60)			(0.61)	(0.77)
Stated SOB					2.16***
					(0.54)
Individual Controls	Yes	Yes	Yes	Yes	Yes
# Observations	876	488	488	656	244
# Participants	122	61	61	82	122

Notes: In the four first columns, the dependent variable is the decision to *Transfer* made for a given induced SOB. In the last column, the dependent variable is the decision to *Transfer* made when the induced SOB corresponded to the stated SOB. Individual controls are: age, gender, guilt-NBE score, fairness score. Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.5: Regression on the decision to Transfer (Logit model with random effects)

Correlation between ...	Risk-Aversion
FOB on Donors' Behavior (Low condition)	0.06
FOB on Donors' Behavior (High condition)	0.09
FOB on Intermediaries' Behavior (Low condition)	0.04
FOB on Intermediaries' Behavior (High condition)	-0.02

N=123; Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.6: Correlation between recipients' beliefs and recipients' risk aversion

	Low condition (χ^2)	High condition (χ^2)
Social Norm on Give	4.17	2.59
Social Norm on Keep	3.61	2.39
Social Norm on Transfer	0.21	1.75
Social Norm on Embezzle	6.89**	0.97

Notes: Kruskal-Wallis tests. Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
[a]: The mean of the intermediaries (0.11) is smaller than the mean of the donors (0.28) (t-test, $p < 0.05$).

Table B.7: Kruskal-Wallis tests of the difference in social norms distributions across roles

C Previous Literature

Study	Game	%	N
Khalmetski et al. (2015)	Dictator	37%	191
Balafoutas and Formwagner (2017)	Dictator	18%	108
Bellemare et al. (2018)	Dictator	$\approx$ 65%	140
Our results	Embezzlement	25%	123

Table C.1: Previous estimations of the proportion of guilt-averse individuals

Study	Game	Estimation	Treatment	θ_i	N
Bellemare et al. (2011)	Proposal and Response	Structural	Dictators' SOB	0.4	1078
			Recipients' FOB	0.8	540
Bellemare et al. (2018)	Dictator	Structural	Stake-independent	0.1	84
			Low Stakes	0.4	56
			Medium Stakes	0.6	56
			High Stakes	1	56
Patel and Smith (2018)	Participation	Equilibrium		0.1	111
Peeters, Vorsatz (2018)	Prisonner Dilemma	Equilibrium	Baseline	2.3	90
			Tempting to coop.	1.8	92
			Tempting to def.	2.5	96
		Hypothetical BDM	Baseline.	3.1	90
			Tempting to coop.	2.1	92
			Tempting to def.	3.5	96
Our results	Embezzlement	Structural	Toward Donor	0.1	61
			Toward Recipient	0.1	62

Table C.2: Previous estimations of the guilt-sensitivity parameter

		Correlation between ...			
Study	Game	Trait	Behaviour	$p < 0.1$	N
Bracht and Regner (2013)	Trust	Guilt-NBE	Pro-social choice	Yes	192
Regner and Harth (2014)	Trust	Moulton's [a]	Pro-social choice	Yes	127
Peeters and Vorsatz (2018)	Prisonner Dilemma	Guilt-NBE	Estimated θ	No	68
Our results	Embezzlement	Guilt-NBE	Pro-social choice	Yes/No	123
			Switching SOB	Yes/No	

Notes: [a] Regner and Harth (2014) used a one question out of the three included in the original measure of Moulton *et al.* (1996): "How easy is it for something to make you feel guilty? (1) very easy, (2) easy, (3) difficult, (4) very difficult".

Table C.3: Previous correlation of personality traits and behavioral outcomes

D Online Appendices

D.1 Instructions for the lab experiment [Translated from French]

OVERVIEW OF THE SESSION

Thank you for participating in this experimental session on decision-making. During this session, you can earn money. The amount of your earnings depends both on your decisions and on other participants' decisions. At the end of the session, you will receive your earnings in cash, in a separate room to ensure the confidentiality of your earnings. The earnings you will receive include:

- your earnings from today's experimental session
- a €7 fee for having completed the online questionnaire and for showing-up on time

During the session, we will sometimes use ECU (Experimental Currency Units). The conversion rate from ECU into Euro is the following: 10 ECU = €1.2.

Please turn off your phone. During the session, any communication with other participants is forbidden. If you have any questions, raise your hand or press the red button on the side of your desk. We will come answer to your questions in private.

At the beginning of the session, the program will form groups of three participants. You will never know the identity of the other two members of your group, and they will never know your identity. All your decisions and earnings are anonymous.

In each group, participants have a different role. There is:

- a donor
- an intermediary
- a recipient

Your screen will indicate your role when the session begins and you will keep the same role throughout the session.

There are two possible situations: situation A and situation B. You will take your decisions in both situations. At every moment, the situation in which you are will always be displayed on the screen.

Short description of the roles

ROLE OF THE DONOR

The donor receives an initial endowment of 150 ECU.

The donor's task is to choose how many ECU to give to the recipient.

For each situation, the donor decides either:

- to give 25 ECU to the recipient
- or to give 0 ECU to the recipient

Regardless of the situation, his/her payoff is equal to: 150 ECU – the ECU given.

Important: The donor cannot give ECU directly to the recipient. Only the intermediary can transfer the ECU given by the donor to the recipient.

ROLE OF THE INTERMEDIARY

The intermediary receives an initial endowment of 80 ECU.

The intermediary's task is to transfer the entirety of the ECU given by the donor to the recipient.

- If the donor has given 25 ECU:

  In situation A, the intermediary can decide either:

  - to transfer the entirety of the 25 ECU to the recipient
  - or to transfer 10 ECU to the recipient and keep 15 ECU for himself/herself

  In situation B, the intermediary can decide either:

  - to transfer the entirety of the 25 ECU to the recipient
  - or to transfer 5 ECU to the recipient and keep 20 ECU for himself/herself

- If the donor has given 0 ECU: The intermediary does not make any decision.

Regardless of the situation, his/her payoff is equal to: 80 ECU + the ECU kept for himself/herself.

Important: For every ECU transferred to the recipient by the intermediary, the recipient receives 2 ECU. For example, if the intermediary transfers 25 ECU, the recipient receives 50 ECU; if the intermediary transfers 5 ECU, the recipient receives 10 ECU.

ROLE OF THE RECIPIENT
The recipient receives an initial endowment of 10 ECU.
The recipient does not make any decision.
Regardless of the situation, his/her payoff is equal to: 10 ECU + (2 x the number of ECU transferred by the intermediary).

Short description of the stages

The session is composed of four stages:

- Stage 1: All the participants answer to some questions.
- Stage 2: The donor makes his/her decisions.
- Stage 3: The intermediary makes his/her decisions.
- Stage 4: All the participants answer to some questions.

At the end of the session:

- All the participants are informed of the randomly selected situation, of the decisions made by the group members in the randomly selected situation, and of their personal earnings.
- All the participants have to complete a final questionnaire.

Personal Login

When I have finished reading these instructions, please enter your personal login on your screen. It corresponds to the personal login you created yourself when you completed the online questionnaire. As a reminder: we advised you to use "Your mother's or father's first name – his/her day of birth – his/her month of birth" without space or dash. If your mother is called Brigitte and she was born on a 19th of May, it yields "Brigitte1905". Once you have entered your personal login, click "Continue".

Comprehension Questionnaire

You have to complete a comprehension questionnaire. If you have any questions, please raise your hand or press the red button. We will come answer to your questions in private.

***

Once all participants have completed the comprehension questionnaire, the session will start. The role that has been randomly assigned to you will be displayed on your screen. You will then receive more detailed instructions.

[The next set of instructions was distributed after the comprehension questionnaire.]

STAGE 1

In this stage, all the participants have to answer to some questions.

If you are an intermediary or a recipient: You will have to answer to the following question: "Among 3 donors randomly selected in today's session, in your opinion how many of these donors will give 25 ECU to the recipient?". You have to enter a number between 0 and 3, inclusive.

You have to answer to this question twice: once in situation A, and once in situation B.

If you are a donor or a recipient: You will have to answer to the following question: "Among 3 intermediaries randomly selected in today's session, if their donor decides to give 25 ECU to the recipient, in your opinion how many of these intermediaries will transfer the 25 ECU to the recipient?".

You have to answer to this question twice: once in situation A, and once in situation B.

In total,

1. If you are a donor, you have to answer to two questions about the intermediaries' decisions (in situation A and in situation B);
2. If you are an intermediary, you have to answer to two questions about the donors' decisions (in situation A and in situation B);
3. If you are a recipient, you have to answer to two questions about the donors' decisions (in situation A and in situation B) and to two questions about the intermediaries' decisions (in situation A and in situation B).

STAGE 2

In this stage, the donors make their decisions.

If you are an intermediary or a recipient, you do not make any decision in this stage.

If you are a donor, your task is to decide whether to give 25 ECU or 0 ECU to the recipient.

In total, you have to make two decisions: one in situation A, and one in situation B. However, only one decision will count to determine the payoff of the group members.

Important: When you make your decisions, you do not know which one of your decision will count. You should give the same weight to each of these decisions since you do not know which one will determine the payoffs of the group members.

Which of the donor's decisions determine the payoffs of the group members?

At the end of the session, the computer program will randomly select situation A or situation B. The donor's decision that will count is the decision that was made in the selected situation.

How does the donor's decision affect the payoffs of the group members?

If the donor has chosen to give 0 ECU to the recipient in the randomly selected situation, the payoff of each group member is the following:

- The donor's payoff is 150 ECU.
- The intermediary's payoff is 80 ECU.
- The recipient's payoff is 10 ECU.

If the donor has chosen to give 25 ECU to the recipient in the randomly selected situation:

- The donor's payoff is 125 ECU.
- - The intermediary's and the recipient's payoffs depend on the intermediary's decisions in the third stage.

At the end of the session, you will be informed of the donor's decision in the randomly selected situation.

***

If you have any question, please raise your hand or press the red button. We will come answer to your questions in private.

[The next set of instructions was distributed after stages 1 and 2.]

STAGE 3

In this stage, the intermediaries make their decisions.

If you are a donor or a recipient, you do not make any decision in this stage.

If you are an intermediary, your task is to transfer the entirety of the ECU given by the donor to the recipient.

You have to make several decisions. Look at the screenshot below. There are two pieces of information in bold characters on the screen: these are the two pieces of information that change for each of the decisions.

Figure D.1: Screenshot for the *Donor Treatment*

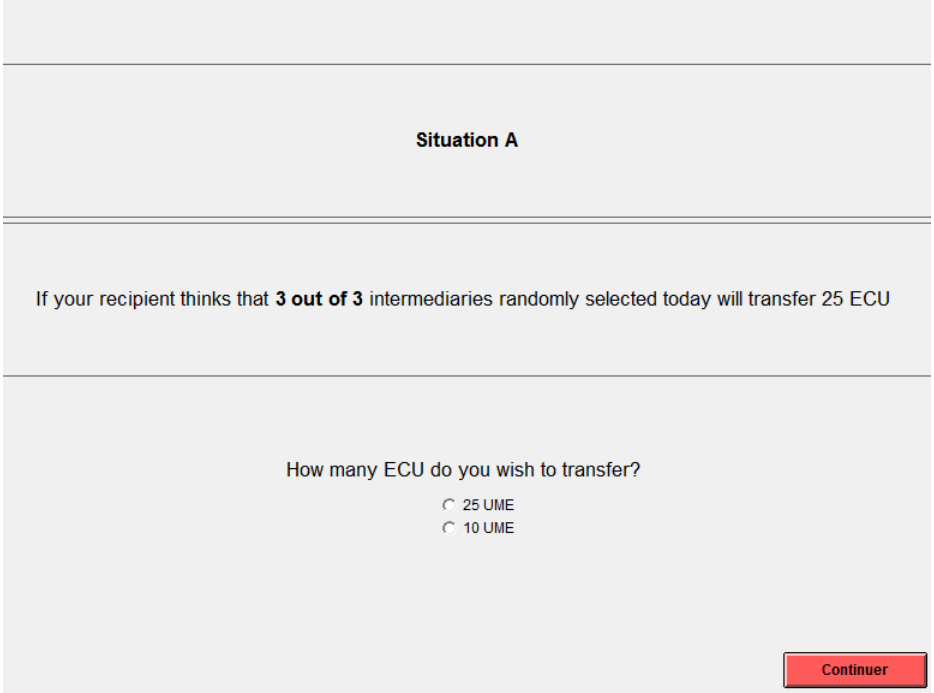

Figure D.2: Screenshot for the *Recipient Treatment*

- **Information on the situation**

You have to make a decision in both situation A and situation B. The order of appearance of these situations on your screen is random.

- **Information on your *donor*/*recipient*'s guess**

Remember that in the first stage the *donor*/*recipient* in your group has answered to the following question: "Among 3 intermediaries randomly selected in today's session, if their donor decides to give 25 ECU to the recipient, in your opinion how many of these intermediaries will transfer the 25 ECU to the recipient?". There were four possible answers: 0, 1, 2 or 3. You have to make a decision for each of the possible answers.

When you make your decisions, you do not know how many ECU the donor in your group has decided to give to the recipient. You have to make your decisions assuming that the donor has given 25 ECU.

In total, you have to make eight decisions: four decisions corresponding to the four possible answers of the *donor*/*recipient* in your group in situation A, and four decisions corresponding to the four possible answers of the *donor*/*recipient* in your group in situation B.

Important: When you make your decisions, you do not know which one of your decision will count. You should give the same weight to each of your decisions since you do not know which one will determine the payoff of the group members.

Which of the intermediary's decisions will determine the payoff of the group members?

- If the donor has chosen to give 0 ECU to the recipient: none of the intermediary's decisions will determine the payoff of the group members.
- If the donor has chosen to give 25 ECU to the recipient: one of the intermediary's decisions will determine the payoff of the group members.

At the end of the session, the computer program will randomly select situation A or situation B. Among the intermediary's decisions made in the randomly selected situation, the computer program selects the decision corresponding to the answer given by the *donor*/*recipient* of your group in the first stage. It is this decision that determines the payoff of the group members.

Example: Suppose that the program randomly selects situation A. Suppose then that, to the question "In situation B, among 3 intermediaries randomly selected in today's session, if their donor decides to give 25 ECU to the recipient, in your opinion how many of these intermediaries will transfer 25 ECU to the recipient?", the *donor*/*recipient* of your group has answered "x". Then, the program selects the decision made by the intermediary when his/her screen displayed "Situation B" and "Your *donor*/*recipient* believes that x intermediaries among 3 randomly selected today will transfer 25 ECU."

How does the intermediary's decision affect the payoff of the group members?

If the donor has given 25 ECU to the recipient in the randomly selected situation, one of the intermediary's decisions determines the payoffs of the group members.

The intermediary may have made three types of decisions:

- Regardless of the situation, if the intermediary transfers 25 ECU to the recipient, the intermediary's payoff is 80 ECU and the recipient's payoff is 60 ECU.
- If situation A is randomly selected and if the intermediary transfers 10 ECU to the recipient and keeps 15 ECU for himself/herself, the intermediary's payoff is 95 ECU and the recipient's payoff is 30 ECU.
- If situation B is randomly selected and if the intermediary transfers 5 ECU to the recipient and keeps 20 ECU for himself/herself, the intermediary's payoff is 100 ECU and the recipient's payoff is 20 ECU.

At the end of the session, you will be informed of the donor's decision in the randomly selected situation.

***

If you have any questions, please raise your hand or press the red button. We will come answer to your questions in private.

[The next set of instructions was distributed after the stage 3]

STAGE 4

1) First, all the participants have to answer to questions of type 1.

You have to evaluate the different possible decisions of a donor and of an intermediary. More precisely, for each possible decision of a donor or of an intermediary, you are asked to indicate whether this decision is socially appropriate and consistent with moral or proper social behavior, or socially inappropriate and inconsistent with moral or proper behavior.

Consider that a decision is socially appropriate if the majority of people agree to say that it is the correct or ethical thing to do. You have to rate each decision using the following scale: very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate or very socially appropriate.

2) Then, the donor and the intermediary have to answer to questions of type 2.

You are asked to guess the decision made by a participant earlier in the session.

How do the answers affect your earnings?

At the end of the session, for each role, the program will randomly select one of the questions to which you have answered in this stage. If you are a recipient, the randomly selected question is for sure a question of type 1. If you are a donor or an intermediary, the question randomly selected can be question of type 1 or a question of type 2.

- If the randomly selected question is a question of type 1:

Your earning depends on the answers of the other participants in the same role as you in today's session. The computer program determines the answer given by the highest number of participants in the same role as you (you included) to this question. You earn 1 if your answer corresponds to the answer the most frequently given by participants in the same role as you. In case of a tie between two answers, the program randomly selects one of the tie answers.

Example: Suppose there are six participants in today's session who have the role of donors. A question of type 1 is randomly selected. To that question, one donor has answered " very socially inappropriate", two donors have answered "somewhat socially appropriate" and three donors have answered "very socially appropriate". The answer the most frequently given by the donors is "very socially appropriate". Then, the three donors who have answered "very socially appropriate" earn 1, the other donors earn nothing.

- If the randomly selected question is a question of type 1:

If you have guessed correctly a previous decision, you earn 1.

END OF THE SESSION

At the end of the session, you will be informed of the situation randomly selected, of the decisions made by your group members in the randomly selected situation, and of your personal payoff. Then, you will be asked to complete a final questionnaire.

At the end of the session, please remain seated and silent until an experimenter invites you to proceed to the payment room. At this moment, bring only your computer tag and your payment receipt completed with you.

***

If you have any questions, please raise your hand or press the red button. We will come answer to your questions in private.

D.2 Online Questionnaire [Translated from French]

PART 0 - Introduction

Thank you for accepting to answer this questionnaire in order to complete your registration to the experiment. Answering to this questionnaire will take approximately 10 minutes. Please read carefully each sentence and remain concentrated. We are interested in your genuine answers, not what you think you should answer.

PART 1 - GASP Questionnaire (Cohen et al., 2011)

Here are situations that people are likely to encounter in day-to-day life, followed by common reactions to those situations. As you read each scenario, try to imagine yourself in that situation.

Please indicate the likelihood that you would react in the way described by using the following categories: (1) Very Unlikely, (2) Unlikely, (3) Slightly Likely, (4) Unlikely, (5) About 50% Likely, (6) Slightly Likely, (7) Very Likely.

1. After realizing you have received too much change at a store, you decide to keep it because the salesclerk does not notice. What is the likelihood that you would feel uncomfortable about keeping the money?
2. You are privately informed that you are the only one in your group that did not make the honor society because you skipped too many days of school. What is the likelihood that this would lead you to become more responsible about attending school?
3. You rip an article out of a journal in the library and take it with you. Your teacher discovers what you did and tells the librarian and your entire class. What is the likelihood that this would make you would feel like a bad person?
4. After making a big mistake on an important project at work in which people were depending on you, your boss criticizes you in front of your co-workers. What is the likelihood that you would feign sickness and leave work?
5. You reveal a friend's secret, though your friend never finds out. What is the likelihood that your failure to keep the secret would lead you to exert extra effort to keep secrets in the future?
6. You give a bad presentation at work. Afterwards your boss tells your co-workers it was your fault that your company lost the contract. What is the likelihood that you would feel incompetent?
7. A friend tells you that you boast a great deal. What is the likelihood that you would stop spending time with that friend?
8. Your home is very messy and unexpected guests knock on your door and invite themselves in. What is the likelihood that you would avoid the guests until they leave?

9. You secretly commit a felony. What is the likelihood that you would feel remorse about breaking the law?

10. You successfully exaggerate your damages in a lawsuit. Months later, your lies are discovered and you are charged with perjury. What is the likelihood that you would think you are a despicable human being?

11. You strongly defend a point of view in a discussion, and though nobody was aware of it, you realize that you were wrong. What is the likelihood that this would make you think more carefully before you speak?

12. You take office supplies home for personal use and are caught by your boss. What is the likelihood that this would lead you to quit your job?

13. You make a mistake at work and find out a co-worker is blamed for the error. Later, your co-worker confronts you about your mistake. What is the likelihood that you would feel like a coward?

14. At a co-worker's housewarming party, you spill red wine on their new cream-colored carpet. You cover the stain with a chair so that nobody notices your mess. What is the likelihood that you would feel that the way you acted was pathetic?

15. While discussing a heated subject with friends, you suddenly realize you are shouting though nobody seems to notice. What is the likelihood that you would try to act more considerately toward your friends?

16. You lie to people but they never find out about it. What is the likelihood that you would feel terrible about the lies you told?

Guilt Negative-Behavior-Evaluation (NBE)	1, 9, 14, 16
Guilt Repair (R)	2, 5, 11, 15
Shame Negative-Self-Evaluation (NSE)	3, 6, 10, 13
Shame Withdraw (W)	4, 7, 8, 12

Table D.1: GASP Questionnaire - Answers Key

PART 2 - Honesty-Humility Scale from the 100-items HEXACO Personality Inventory - Revised (Lee and Ashton, 2004)

Please indicate how much you agree or disagree with these statements about you by using the following categories: (1) Strongly disagree, (2) Disagree, (3) Neutral (neither agree nor disagree), (4) Agree, (5) Strongly disagree.

1. If I want something from a person I dislike, I will act very nicely toward that person in order to get it.
2. If I knew that I could never get caught, I would be willing to steal a million dollars.
3. Having a lot of money is not especially important to me.
4. I am an ordinary person who is no better than others are.
5. I would not use flattery to get a raise or promotion at work, even if I thought it would succeed.
6. I would be tempted to buy stolen property if I were financially tight.
7. I would like to live in a very expensive, high-class neighborhood.
8. I would not want people to treat me as though I were superior to them.
9. If I want something from someone, I will laugh at that person's worst jokes.
10. I would never accept a bribe, even if it were very large.
11. I would like to be seen driving around in a very expensive car.
12. I think that I am entitled to more respect than the average person is.
13. I would not pretend to like someone just to get that person to do favors for me.
14. I would be tempted to use counterfeit money, if I were sure I could get away with it.
15. I would get a lot of pleasure from owning expensive luxury goods.
16. I want people to know that I am an important person of high status.

Sincerity	1R, 5, 9R, 13
Fairness	2R, 6R, 10, 14R
Greed-Avoidance	3, 7R, 11R, 15R
Modesty	4, 8, 12R, 16R

Table D.2: Honesty-Humility Scale - Answers Key[26]

PART 3 – Inspired by the Self Report Altruism Scale (Rushton et al., 1981)[27]

Please indicate the frequency with which you have carried out the following acts by using the following categories: (1) Never, (2) Once, (3) More than once, (4) Often, (5) Very Often.

[27]Three items were excluded: "I have made change for a stranger", "I have given a stranger a lift in my car" and "I have bought 'charity" Christmas cards deliberately because I knew it was a good cause".

1. I have helped a stranger change a flat tire.[28]
2. I have given directions to a stranger.
3. I have given money, goods or clothes to a charity.[29]
4. I have delayed an elevator and held the door open for a stranger.
5. I have donated blood.
6. I have helped carry a stranger's belongings (books, parcels, etc.).
7. I have allowed someone to go ahead of me in a lineup (at photocopy machine, in the supermarket).
8. I have pointed out a clerk's error (in a bank, at the supermarket) in undercharging me for an item.
9. I have let a neighbor whom I did not know too well borrow an item of some value to me (e.g., a dish, tools, etc.)
10. I have done volunteer work for a charity.
11. I have helped a classmate who I did not know that well with a homework assignment when my knowledge was greater than his or hers.
12. I have before being asked, voluntarily looked after a neighbor's pets or children without being paid for it.
13. I have offered to help a handicapped or elderly stranger across a street.
14. I have offered my seat on a bus or train to a stranger who was standing.
15. I have helped an acquaintance to move households.
16. I have given money to a stranger who needed it (or asked me for it).

PART 4 – Socio-Demographics

1. Risk Preferences (Dohmen et al., 2011)
   How would you describe yourself? Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please tick a box on the scale, where the value "0" means "not at all willing to take risks" and the value "10" means "very willing to take risks".

[28] Originally: "I have helped push a stranger's car out of the snow."
[29] Originally it was two different items: I have given money to charity" and "I have donated goods or clothes to a charity".

2. Time Preferences (Visher et al., 2013)
   How would you describe yourself? Are you generally an impatient person, or someone who always shows great patience? Please tick a box on the scale, where the value "0" means "very impatient" and the value "10" means "very patient".

3. Religiosity
   How would you describe yourself? How often do you pray?

   - I never pray
   - I seldom pray
   - I pray every week
   - I pray more than once a day

4. Gender
   Please indicate your gender.

   - Female
   - Male

5. Age
   Please indicate your age.

6. Status
   Please indicate your status.

   - Student
   - Employed
   - Unemployed
   - Retired

   (a) School - *if your answer to question 6 is "Student"*
   Which school do you attend?
   - EM Lyon
   - Ecole Centrale Lyon
   - ISOstéo
   - Université Lyon 1
   - Université Lyon 2
   - Université Lyon 3
   - Université Catholique de Lyon
   - Other

   (b) Field of Study - *if your answer to question 6 is "Student"*
   What is your field of study?

- Economics and Management
- Social Sciences
- Arts and Humanities
- Engineering Sciences
- Medical Studies
- Other

(c) Professional Activity - *if your answer to question 6 is "Employed"*
What is your current professional status?

- Farmer
- Craftsman, shopkeeper, business owner
- Executive and higher intellectual occupations
- Civil servant, administrative employee
- Employee
- Worker

7. Number of previous experiments
In how many GATE-LAB experimental sessions have you participated already?

8. Personal Login
Please choose a personal login. Choose a login that you can remember easily since you will need this login to start the experimental session. We suggest you use "Mother's or Father's first name - her/his day of birth - her/his month of birth" without space or dash. For example, if your mother is called Brigitte and is born a May 19th, the suggested login is "Brigitte1905".