DISCUSSION PAPER SERIES

Embezzlement and Guilt Aversion

Giuseppe Attanasi
Claire Rimbaud
Marie Claire Villeval

DISCUSSION PAPER SERIES

Embezzlement and Guilt Aversion

Giuseppe Attanasi
University of Nice-Sophia Antipolis

Claire Rimbaud
University of Lyon

Marie Claire Villeval
University of Lyon

NOVEMBER 2018

ABSTRACT

Embezzlement and Guilt Aversion[*]

Psychological game theory can contribute to renew the analysis of unethical behavior by providing insights on the nature of the moral costs of dishonesty. We investigate the moral costs of embezzlement in situations where donors need intermediaries to transfer their donations to recipients and where donations can be embezzled before they reach the recipients. We design a novel three-player Embezzlement Mini-Game to study whether intermediaries in the laboratory suffer from guilt aversion and whether guilt aversion affects the decision to embezzle. We show that the proportion of guilt-averse intermediaries is the same irrespective of the direction of guilt and guilt aversion reduces embezzlement. Structural estimates indicate no difference in the effect of guilt aversion toward the donor and toward the recipient on intermediaries' behavior. This is striking as embezzlement affects the earnings of the recipient but not those of the donor. It shows that guilt aversion matters even when decisions have no direct monetary consequences.

Corresponding author:
Marie Claire Villeval
Université de Lyon
CNRS
GATE UMR 5824
93 Chemin des Mouilles
F-69130 Ecully
France
E-mail: villeval@gate.cnrs.fr

1 Introduction

Contrary to the standard economics-of-crime approach that focuses on the trade-off between the monetary costs and benefits of dishonesty (Becker and Stigler, 1974; Fan *et al.*, 2009), the behavioral economic analysis of unethical behavior insists on the importance of incorporating moral costs in this trade-off. Indeed, not all individuals cheat, even when there is no risk of detection, and most cheaters do not exploit their opportunity of cheating maximally, which may come from the moral costs associated with unethical behavior (see, *e.g.*, Gneezy, 2005; Mazar *et al.*, 2008; Fischbacher and Föllmi-Heusi, 2013; Kajackaite and Gneezy, 2017; Abeler *et al.*, 2018, in the context of lying, and Abbink and Serra, 2012; Drugov *et al.*, 2014; Köbis *et al.*, 2016, in the context of corruption). However, little is known on the nature of these moral costs beyond the idea that most people are willing to maintain a positive self-image. Psychological game theory, introduced by Geanakoplos *et al.* (1989) and further developed by Battigalli and Dufwenberg (2009), helps to understand the nature of these moral costs through the modeling of emotions such as guilt aversion, although this theory has been rarely used so far to investigate dishonesty (for a recent exception, see Dufwenberg and Dufwenberg, 2018).

In this paper, we study guilt aversion in the context of embezzlement. Embezzlement is defined as the misappropriation of assets by individuals to whom they were entrusted, in order to monopolize or to steal them. It can be observed when the providers of resources need intermediaries to transfer these resources to the final recipients. The problem is crucial, especially in developing countries, in domains such as health, education, or humanitarian aid where the final recipients seldom receive the totality of aid transfers they are entitled to.[1] Indeed, donors must rely on local intermediaries and usually cannot verify which amount has eventually been transferred to the entitled recipients. Embezzlement is detrimental to economic development and

---

[1]For example, in 2000 in Ghana, a Public Expenditure Tracking Survey revealed that 80% of non-salary funds did not reach health facilities (Canagarajah and Ye, 2001). For the period 1991-1995, Ugandan schools received on average 13% of the governmental transfers they were entitled to (Reinikka and Svensson, 2004). 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.

cooperation (Olken and Pande, 2012; Beekman *et al.*, 2014) and it can result in some programs becoming inequality enhancing (*e.g.*, Reinikka and Svensson, 2004) or no longer cost-effective (*e.g.*, Ferraz *et al.*, 2012).

While most of the previous literature has studied interventions affecting the monetary costs and benefits attached to embezzlement (*e.g.*, Di Tella and Schargrodsky, 2003; Olken, 2007; Barr *et al.*, 2009), we investigate the moral 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). Our first research objective is to identify in the laboratory the existence of such guilt aversion and its impact on the behavior of intermediaries who can embezzle the donations made by donors to recipients. Our second objective is to test whether the *direction of guilt aversion* matters, *i.e.*, whether it is stronger toward the donor or toward the recipient.

We designed a novel three-player game – the Embezzlement Mini-Game. In this game, a donor sends a donation to a recipient but this donation has to be transferred by an intermediary who can embezzle a fraction of this donation to increase his own material payoff.[2] Embezzlement decreases both the utility of the donor who cares about the recipient's well-being and the utility of the recipient who receives the donation. The donor forms expectations on how much of the donation the intermediary will transfer to the recipient. 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, the intermediary may be affected by *donor-guilt aversion* and by *recipient-guilt aversion*.

Indirect evidence of intermediaries' guilt aversion can be found in previous studies. Chlaß *et*

---

[2]The game is meant to represent a situation in which an individual in a rich country sends money to a charity to help improve the situation of individuals in need in developing countries. The donor and the charity have to rely on local intermediaries, *e.g.*, the heads of villages. In many cases, these intermediaries are in a position to embezzle part of the donations – see for example the field experiments reported in Beekman *et al.* (2014).

al. (2015) found that the more intermediaries believe that donors have donated, the more they transfer. This is coherent with our model's intuition which predicts that the more donors believe the donation will be transferred, the more intermediaries transfer. Di Falco *et al.* (2016) found that intermediaries at the beginning of longer transfer chains embezzle less than intermediaries in short chains. Feeling guilty from letting down the recipients' expectation could explain this behavior.

We rely on the modeling of simple 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. This theory departs from traditional game theory in assuming that players' utilities do not only depend on their decisions but also on their beliefs about decisions, beliefs, or information. In particular, the psychological utility of a guilt-averse player depends on his second-order beliefs, *i.e.*, his beliefs about other players' beliefs about his own decision. For the recipient-guilt aversion of the intermediary, we rely on BD's (2007) definition of guilt as the disutility from letting down the other player's expectations about *his own* payoff. For the donor-guilt aversion of the intermediary, we extend theoretically BD (2007) model of simple guilt by introducing a novelty in the definition of guilt in the psy-games literature. Rather than not letting down the donor's expectations about his own payoff – which is not affected by the decision to embezzle –, a donor-guilt averse intermediary dislikes letting down the donor's expectations about *another player*'s material payoff, *i.e.*, the recipient. In this case, the psychological utility of the guilt-averse player (the intermediary) depends on his beliefs about another player's beliefs (the donor) on a *third* player's material payoff (the recipient).[3]

Our theoretical analysis builds on the incomplete-information framework with role-dependent guilt of Attanasi *et al.* (2016). We assume that among the two active players only the intermediary can feel guilty. We enrich the set of psychological types by assuming that both the donor and

---

[3]In Balafoutas (2011), an extension of his dynamic model also allows for an official who accepts a bribe to feel guilty toward both the citizen and the lobby. However, both directions of guilt are coherent with BD's (2007) model since the official can affect the payoff of both the citizen and the lobby.

the intermediary have altruistic preferences toward the recipient. Unlike Attanasi *et al.* (2016), we elaborate our behavioral hypotheses relying on best-reply analysis rather than on Bayesian equilibrium. This is motivated by the fact that a standard equilibrium analysis has no compelling foundation for games played one-shot (like ours) and in experiments on other-regarding preferences. Furthermore, and more importantly, in a psychological type space with the donor's and intermediary's altruism toward the recipient and with the intermediary's guilt toward the recipient and the donor, best-reply analysis can be carried out regardless of (in)completeness of information about players' types.[4] Thus, it delivers sharp predictions on the correlation between the intermediary's guilt types and behavior, and between his second-order beliefs and behavior, independently of the direction of guilt aversion (donor-guilt or recipient-guilt). Predicting the sign and size of these correlations is enough to provide appropriate behavioral hypotheses given the two research objectives mentioned above.

We implemented 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). Between-subjects, we manipulated the information given to the intermediaries before they made their decision. In the *Donor treatment*, intermediaries decided whether to transfer 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. Within-subjects, we manipulated the percentage of the donation that could be embezzled (80% in the High condition and 60% in the Low condition) to test how the intensity of potential embezzlement affects beliefs.

Our results 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

---

[4]Indeed, ours is an incomplete-information framework with private values. Hence, beliefs about the types of others do not enter the best-reply correspondence.

the guilt and of the percentage of the donation that could be embezzled. Structural estimates indicate no difference in the effect of guilt aversion toward the donor and toward the recipient on intermediaries' behavior. This shows that guilt aversion may influence behavior even when decisions have no direct monetary consequences on the person toward whom guilt is directed.

The remainder of this paper is organized as follows. Section 2 introduces the theoretical model and its predictions. Section 3 describes the experimental design. Section 4 presents the 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.
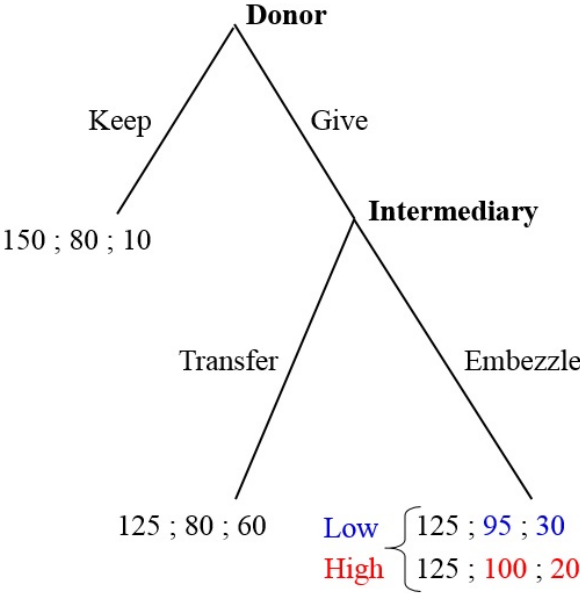
Figure 1: The Embezzlement Mini-Game(s)

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 the median between the donor's and the recipient's endowments.[5]

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 receives twice the amount actually transferred. Thus, embezzlement involves an efficiency loss.[6]

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 $f \in \{0, 0.8\}$).

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 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

Figure 1 shows respectively the *Donor*'s, *Intermediary*'s and *Recipient*'s material payoff (M_j, with $j \in \{D, I, R\}$) at each terminal node of the Embezzlement Mini-Game, *i.e.*, for each profile of donor's and intermediary's strategy, respectively $s_D \in \{Keep, Give\}$ and $s_I \in \{Transfer$ if $Give$,

---

[5]The intermediary can be seen as the middleman in a network linking a NGO or a Governmental Agency to villagers. Different sets of possible actions for the donor and the intermediary capture asymmetry of positions. Unlike in a consecutive three-person dictator game (Bahr and Requate, 2013), the different initial endowments underline the different status of each player.

[6]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 should reinforce the immoral image of embezzlement.

Embezzle if *Give*}. We denote the donor's strategies *Keep* and *Give* with respectively K and G, and the intermediary's strategies $Transfer$ if *Give* and *Embezzle* if *Give* with respectively T and E.

First of all, we assume that the **recipient's utility function** coincides with his material payoff, *i.e.*, $U_R(s_D, s_I) = M_R(s_D, s_I)$, which is made of his initial endowment, and the amount received $r(s_D, s_I)$. The latter enters the recipient's utility function only if the donor chooses *Give*, *i.e.*, $s_D = G$. In that case, the amount received depends on the amount actually transferred by the intermediary, $r(s_I)$. We are interested in the recipient's beliefs only in terms of their psychological impact on the intermediary's strategy s_I. 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*. Hence, our focus is on the recipient's expected received amount in this subgame:

$$\mathbb{E}_R[r(s_I)|s_D = G] = \alpha_{RI} \cdot r(T) + (1 - \alpha_{RI}) \cdot r(E) \tag{1}$$

Let us now introduce the **donor's utility function**. It is composed of his material payoff and his feeling of altruism toward the recipient (Eq. (2)). We assume that the donor (as well as the intermediary) have altruistic preferences toward the recipient. A player's feeling of altruism, A_{jR} with $j \in \{D, I\}$, represents player j's utility derived from an increase in the amount received by the recipient (*belief-independent* preferences). It is the product of two terms: $\gamma_j \geq 0$, player j's altruism sensitivity toward the recipient, *i.e.*, his altruistic type, and $r(s_D, s_I)$, the amount actually received by the recipient. With this, the donor's utility is:

$$U_D(\gamma_D, s_D, s_I) = M_D(s_D) + A_{DR}(\gamma_D, s_D, s_I) \tag{2}$$

$$\text{where } A_{DR}(\gamma_D, s_D, s_I) = \gamma_D \cdot r(s_D, s_I) \tag{3}$$

When he chooses between *Keep* and *Give*, he does not know what would be the intermediary's strategy. Therefore, his first-order belief that the intermediary chooses *Transfer* after *Give*, α_{DI},

matters for his giving choice. His expected utility conditional on choosing *Give* is:

$$\mathbb{E}_D[U_D(\gamma_D, s_I)|s_D = G] \quad = \quad M_D(G) + \gamma_D \cdot \mathbb{E}_D[r(s_I)|s_D = G] \tag{4}$$

where the amount the donor expects the recipient to get after his *Give* choice is:

$$\mathbb{E}_D[r(s_I)|s_D = G] = \alpha_{DI} \cdot r(T) + (1 - \alpha_{DI}) \cdot r(E) \tag{5}$$

We made the simplifying assumption that a donor prefers that his donation increases the recipient's payoff rather than the intermediary's. This is broadly consistent with other models of distributional preferences: inequity aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), since the recipient is the most disadvantaged player, and concern for efficiency (Charness and Rabin, 2002), since the sum of payoffs is maximized if the donor Gives and the intermediary Transfers. Importantly, 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, we should find that the frequency of giving increases in the donor's first-order belief about *Transfer* after *Give*.

We finally introduce the **intermediary's utility function**. The intermediary's utility (Eq. (6)) is composed of his material payoff M_I, his feeling of altruism toward the recipient A_{IR} (Eq. (7)), and his feeling of guilt toward the other players B_{Ij}, with $j \in \{D, R\}$ (Eq. (8)). As anticipated, we assume that the intermediary has altruistic preferences toward the recipient and that they are modeled as the donor's ones (see Eqs. (2) and (3)). Second, in line with the role-dependent guilt model of Attanasi *et al.* (2016), we assume that only the intermediary can feel guilty.[7] The intermediary's feeling of guilt, B_{Ij}, with $j \in \{D, R\}$, represents his disutility

---

[7] See the discussion in Attanasi *et al.* (2016), p. 649, where they argue that role dependence of guilt preferences is plausible in asymmetric games (see, *e.g.*, Attanasi *et al.*, 2013, 2018, for indirect experimental evidence corroborating the assumption). In particular, they discuss how the assumption that sensitivity to guilt is triggered only when playing in the role of trustee (and not in the role of trustor) in the Trust Game resonates with the evolutionary psychology of emotions and the conceptual act theory

derived from letting down other players' expectations on the strategy he will select (*belief-dependent* preferences). It is the product of two terms: $\theta_{Ij} \geq 0$, the guilt sensitivity toward player $j \in \{D, R\}$, *i.e.*, the intermediary's guilt type; and the difference, if positive, between player j's expectations on the transferred amount after $Give$, $\mathbb{E}_j[r(s_I)|s_D = G]$, and the amount actually transferred to the recipient $r(s_I)$. This difference depends both on the intermediary's strategy, and on player j's first-order belief about this strategy (see Eqs. (1) and (5), respectively for $j = R$ and $j = D$). If $\mathbb{E}_j[r(s_I)|s_D = G] > r(s_I)$, then the intermediary feels guilty from letting down player j's expectations on the amount transferred to the recipient. Independently from the treatment, the intermediary does not feel guilty if $(s_D, s_I) = (G, T)$, *i.e.*, the donor gives and the intermediary transfers the whole donation to the recipient. With this, the intermediary's utility is, for $j \in \{D, R\}$:

$$U_I(\theta_{Ij}, \gamma_I, s_I, \alpha_{jI}|s_D = G) = M_I(G, s_I) + A_{IR}(\gamma_I, s_I) - B_{Ij}(\theta_{Ij}, s_I, \alpha_{jI}) \tag{6}$$

$$\text{where } A_{IR}(\gamma_I, s_I) = \gamma_I \cdot r(s_I) \tag{7}$$

$$\text{and } B_{Ij}(\theta_{Ij}, s_I, \alpha_{jI}) = \theta_{Ij} \cdot \max\{0, \mathbb{E}_j[r(s_I)|s_D = G] - r(s_I)\} \tag{8}$$

Two clarifications are in order about Equation (8). First, we analyze the impact of each guilt sensitivity (toward the donor and toward the recipient) separately because of our experimental design. We use 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. Therefore, we make the auxiliary assumption that one direction of guilt prevails over the other in each treatment, *i.e.*, $\theta_{IR} = 0$ in the Donor treatment and $\theta_{ID} = 0$ in the Recipient treatment.

Second, for B_{IR} (Eq. (8) with $j = R$), we rely on BD's (2007) definition of simple guilt as the intermediary's disutility from letting down the recipient's expectations about *his own*

---

of emotion. Similar arguments can be provided in support of sensitivity to guilt being triggered only when playing in the role of intermediary (and not in the role of donor) in the Embezzlement Mini-Game.

material payoff, whereas, for B_{ID} (Eq. (8) with $j = D$), we extend BD (2007) by defining the intermediary's guilt as the disutility from letting down the donor's expectations about *the recipient*'s material payoff ($\mathbb{E}_D[r(s_I)|s_D = G]$).

2.3   Theoretical Predictions

We provide a best-reply analysis of the Embezzlement Mini-Game(s) with incomplete information. We assume for simplicity that the recipient is commonly known to be selfish. But neither the donor's altruistic type, γ_D, nor the intermediary's guilt-altruistic type, (θ_{Ij}, γ_I), are known to the co-players. The analysis relies on the assumption of players' **Rationality**: each player is rational, *i.e.*, a subjective expected utility maximizer.[8]

2.3.1   Predictions on Donor's behavior

We define the donor's *Willingness-to-Give function* (*WG*) as the difference between his (expected) utility from *Give* (Eq. (5)) and his (certain) utility from *Keep*, the latter coinciding with his initial endowment, $M_D(K)$. The more the donor prefers to *Give* rather than *Keep*, the higher his willingness to *Give*:

$$
\begin{aligned}
WG(\gamma_D, \alpha_{DI}) &= \mathbb{E}_D[U_D(\gamma_D, s_I | s_D = G] - U_D(\gamma_D | s_D = K) \\
&= M_D(G) - M_D(K) + \gamma_D \cdot (\alpha_{DI} \cdot r(T) + (1 - \alpha_{DI}) \cdot r(E)) \quad (9)
\end{aligned}
$$

Under the two parametrizations of the Mini-Embezzlement Game in Figure 1, a rational donor prefers to *Give* in the Low condition if $WG = -25 + \gamma_D \cdot (\alpha_{DI} \cdot 30 + 20) > 0$, and in the High condition if $WG = -25 + \gamma_D \cdot (\alpha_{DI} \cdot 40 + 10) > 0$. This leads to the following set of 'type-belief' pairs consistent with a *R*ational donor choosing *Give* in the Low and High conditions,

---

[8] A two-step rationalizability procedure based on forward-induction reasoning (cf. BD, 2009, Section 5; Battigalli *et al.*, 2018a, 2018b) would provide similar qualitative predictions, by assuming that $\theta_{ID} > 0$ and $\theta_{IR} > 0$ at the same time in both treatments. Technical details of this analysis are available from the authors upon request.

respectively:

$$R_D^{G|Low} = \left\{ (\gamma_D, \alpha_{DI}) : \alpha_{DI} \geq \frac{1}{6} \left(\frac{5}{\gamma_D} - 4 \right) \right\} \tag{10}$$

$$R_D^{G|High} = \left\{ (\gamma_D, \alpha_{DI}) : \alpha_{DI} \geq \frac{1}{8} \left(\frac{5}{\gamma_D} - 2 \right) \right\} \tag{11}$$

Eqs. (10) and (11) are represented in Figure 2. The figure shows the $(s_D|Low, s_D|High)$ regions of the donor's 'type-belief' space (γ_D, α_{DI}), where the rational donor is predicted to: *Keep* in both conditions (white region); *Give* in the Low and *Keep* in the High condition (light grey region); *Give* in both conditions (dark grey region). From Eqs. (10) and (11), it is easy to check that $(R_D^{K|Low} \cap R_D^{G|High}) = \varnothing$, *i.e.*, holding the 'type-belief' pair constant across conditions, a *R*ational donor cannot choose *Keep* in the Low and *Give* in the High condition, which explains the absence of a $(Keep, Give)$ region in Figure 2. The horizontal lines indicate the four possible first-order beliefs about the intermediary that a donor can hold in our experiment, $\alpha_{DI} \in \{0, 1/3, 2/3, 1\}$, as we will explain in Section 3.1.
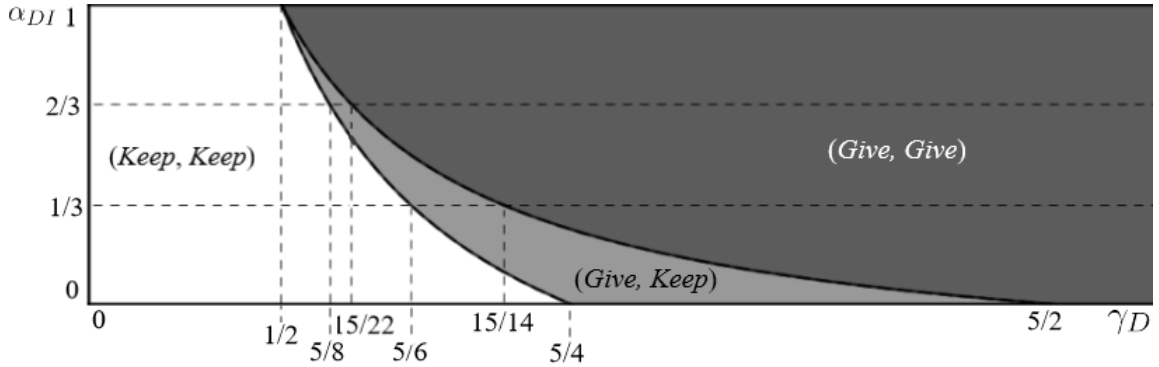
Figure 2: Predicted behavior of a rational donor in the two conditions (Low, High), depending on his altruistic type (γ_D) and first-order belief (α_{DI})

A comparative static analysis across the three regions of predictions in Figure 2 allows us to elaborate our hypotheses about the donor's behavior.

First, let us fix the pair $(\gamma_D, \alpha_{DI} = 0)$, *i.e.*, a donor with no trust on the intermediary's *Transfer* choice, and let us increase his first-order belief α_{DI}. For $\gamma_D \in [0, 1/2]$, the donor prefers

11

to *Keep* in both conditions, for any α_{DI} (white region). For $\gamma_D \in (1/2, 5/4]$, as α_{DI} begins to increase, the donor switches from *Keep* to *Give* in the Low condition (light grey region); if α_{DI} continues to increase, he switches from *Keep* to *Give* also in the High condition (dark grey region). For $\gamma_D \in (5/4, 5/2]$, the donor prefers to *Give* in the Low condition, for any α_{DI} (light grey region); as α_{DI} increases, the donor switches from *Keep* to *Give* in the High condition (dark grey region). For $\gamma_D \in (5/2, +\infty)$, the donor prefers to *Give* in both conditions, for any α_{DI} (dark grey region). Therefore, independently from the condition, an increase in α_{DI} never leads to a switch from *Give* to *Keep* and for some subset of donor's sensitivities to altruism it leads to a switch from *Keep* to *Give*. Considering heterogeneity in donors' types, we elaborate a hypothesis about the donor's **belief-dependent behavior**, whose verification is crucial to validate our assumption that the donor's utility increases with the recipient's payoff.

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

Now suppose that the pair (γ_D, α_{DI}) is the same in the Low and High conditions, and refer again to Figure 2. If this 'type-belief' pair belongs to the white or the dark grey region, the rational choice is the same in both conditions, while if it lies in the light grey region, the rational choice is *Give* in the Low and *Keep* in the High condition. There is no 'type-belief' pair in the light grey region for $\alpha_{DI} = 1$, *i.e.*, when the donor is certain that the intermediary will *Transfer*. In that case, being the payoff profile after history (*Give*, *Transfer*) invariant to the condition (see Figure 1), the donor's WG in Eq. (9) is the same both in the Low and High condition, and so the predicted choice. Since conditions are manipulated within-subjects (see Section 3.1), we assume that the distribution of donors' types is the same across conditions. Belief elicitation in the two conditions will allow us to check their invariance to the condition, that we assume in order to elaborate a hypothesis about the donor's **condition-dependent behavior**.

H.D2 [High *vs.* Low Condition on Choice]: Given a donor's first-order belief about *Transfer*

lower than one, same in both conditions, the frequency of *Give* choices by altruistic donors is higher in the Low than in the High condition.

Finally, let us fix the pair $(\gamma_D = 0, \alpha_{DI})$, *i.e.*, a selfish donor. Figure 2 shows that as his sensitivity to altruism γ_D increases, the donor moves from the white region directly to the dark grey region for $\alpha_{DI} = 1$, or passing through the light grey region for all $\alpha_{DI} < 1$. Therefore, independently from the condition, an increase in γ_D never leads to a switch from *Give* to *Keep* and it can lead to a switch from *Keep* to *Give*. Considering heterogeneity in donors' sensitivity to altruism, we elaborate a hypothesis about the donor's **type-dependent behavior**.

H.D3 [Choice-Type Correlation]: For a given first-order belief about *Transfer*, the frequency of *Give* choices increases with the donor's sensitivity to altruism.

Note that we derived H.D1, H.D2, and H.D3 without specifying the treatment (Donor or Recipient) since, in our experiment, donors are unaware of the treatment when they make their choices. Therefore, the donor's behavior should be **treatment-independent**.

2.3.2   Predictions on Intermediary's behavior

Relying on Eqs. (6–8), we define for each treatment (Donor and Recipient) the intermediary's *Willingness-to-Transfer function* (*WT*) as the difference between his utility when he Transfers and his utility when he Embezzles. Both these terms are expected utilities since the intermediary forms beliefs about the first-order beliefs α_{jI} of the co-player j toward whom he feels guilty ($j = D$ in the Donor and $j = R$ in the Recipient treatment).[9] These are his conditional second-order beliefs $\beta_{Ij} = \mathbb{E}_I[\alpha_{jI}|s_D = G]$, *i.e.*, for $j \in D, R$, conditional on the donor choosing *Give*.[10]

---

[9]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*. Indeed, in our experiment 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* (see Section 3.1).

[10]More precisely, we reason as if the intermediary has a point belief β_{Ij} about α_{jI} conditional on *Give*.

The more the intermediary prefers to *Transfer* rather than *Embezzle*, the higher his willingness to *Transfer*.[11] Thus, for $j \in \{D, R\}$:

$$
\begin{aligned}
WT(\theta_{Ij}, \gamma_I, \beta_{jI} | s_D = G) &= \mathbb{E}_I[U_I(\theta_{Ij}, \gamma_I, \alpha_{jI} | G, T)] - \mathbb{E}_I[U_I(\theta_{Ij}, \gamma_I, \alpha_{jI} | G, E)] \\
&= M_I(G, T) - M_I(G, E) + \gamma_I \cdot [r(T)) - r(E)] + \\
&\quad\; \theta_{Ij} \cdot [\beta_{Ij} \cdot r(T) + (1 - \beta_{Ij}) \cdot r(E)) - r(E)]
\end{aligned}
\tag{12}
$$

Rationality of the intermediary implies, for $j \in \{D, R\}$, that type (θ_{Ij}, γ_I) with belief β_{Ij} chooses to *Transfer* the donation if $WT > 0$ in Eq. (12) and to *Embezzle* a fraction of it otherwise. Under the two parametrizations of the Mini-Embezzlement Game in Figure 1, Eq. (12) becomes, for $j \in \{D, R\}$, $WT = -15 + 30 \cdot (\gamma_I + \theta_{Ij} \cdot \beta_{Ij})$ in the Low condition and $WT = -20 + 40 \cdot (\gamma_I + \theta_{Ij} \cdot \beta_{Ij})$ in the High condition. This leads to the following set of 'type-belief' pairs consistent with a *Rational* intermediary choosing *Transfer* in the Low and High conditions:

$$
R_I^{T|Low} = R_I^{T|High} = \left\{ ((\theta_{Ij}, \gamma_I), \beta_{Ij}) : \gamma_I + \theta_{Ij} \cdot \beta_{Ij} > \frac{1}{2} \right\}
\tag{13}
$$

Note that for each 'type-belief' pair $((\theta_{Ij}, \gamma_I), \beta_{Ij})$, the sign of WT is the same in both conditions. Thus, also the complementary set of 'type-belief' pairs consistent with a *Rational* intermediary choosing *Embezzle* is independent from the condition, *i.e.*, $R_I^{E|Low} = R_I^{E|High}$.

Figure 3 shows the regions of the intermediary's type space (θ_{Ij}, γ_I), where a rational intermediary is predicted to *Keep* in both conditions or *Give* in both conditions for fixed conditional second-order beliefs about *Transfer*.[12] More precisely, for each type (θ_{Ij}, γ_I) it is shown the best-reply strategy for $\beta_{Ij} \in \{0, 1/3, 2/3, 1\}$. The four dotted lines indicate types indifferent between *Embezzle* and *Transfer* for each of these β_{Ij}. Thus, *e.g.*, types in the white region *Embezzle* for

---

[11]Note that if the intermediary Transfers he experiences no guilt, and so $B_{Ij} = 0$ in Eq. (8).

[12]Recall that, as anticipated above, in our experiment both the donor and the recipient can only hold four possible first-order beliefs about *Transfer*, $\alpha_{jI} \in \{0, 1/3, 2/3, 1\}$ (see Section 3.1). With this, we make the operational assumption that also the intermediary can only hold four possible second-order beliefs, $\beta_{Ij} \in \{0, 1/3, 2/3, 1\}$, on each co-player $j \in \{D, R\}$.

all the four possible β_{Ij}, while types in the lightest-grey region *Embezzle* for $\beta_{Ij} \in \{0, 1/3, 2/3\}$ and *Transfer* for $\beta_{Ij} = 1$, *i.e.*, they switch from *Embezzle* to *Transfer* for $\beta_{Ij} = 1$.
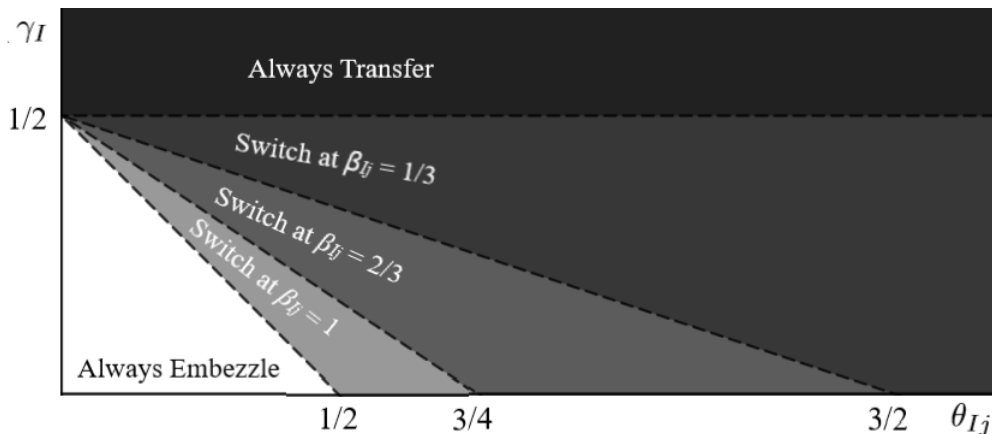
Figure 3: Predicted behavior of a rational intermediary for the four possible second-order beliefs $\beta_{Ij} \in \{0, 1/3, 2/3, 1\}$, depending on his guilt type (θ_{Ij}) and altruistic type (γ_I)

A comparative static analysis of WT in Eq. (12) and of the four regions of predictions in Figure 3 allows us to elaborate our hypotheses about the intermediary's behavior.

First of all, Figure 3 shows that if $\gamma_I > 1/2$, then the intermediary always Transfers, independently from θ_{Ij} and β_{Ij}. In that case, sensitivity to altruism is sufficiently high to prevail over guilt aversion. For $\gamma_I < 1/2$, if the intermediary is guilt-averse ($\theta_{Ij} > 0$), WT in Eq. (12) is increasing in β_{Ij}, *i.e.*, $\frac{\delta WT}{\delta \beta_{Ij}} > 0$ independently from the condition Low or High. Therefore, our first hypothesis is about the intermediary's **belief-dependent behavior**:

H.I1 [Choice-Belief Correlation]: For sufficiently low sensitivity to altruism toward the recipient, the frequency of *Transfer* choices by guilt-averse intermediaries increases in their second-order beliefs about *Transfer*. Intermediaries with sufficiently high sensitivity to altruism choose to *Transfer* regardless of their second-order beliefs.

The second part of H.I1 suggests that the fraction of guilt-averse intermediaries in the sample of experimental participants might be underestimated by only looking at their behavior. In fact,

15

although donor-guilt or recipient-guilt averse, they could disclose a belief-independent *Transfer* pattern due to a sufficiently high sensitivity to altruism toward the recipient.

Furthermore, knowing from Eq. (13) that $R_I^{s_I|Low} = R_I^{s_I|High}$ for $s_I \in \{T, E\}$, we can elaborate the following hypothesis about the intermediary's **condition-independent behavior**. Note that conditions are manipulated within-subjects, thus we can assume that the distribution of the intermediaries' types is the same across conditions. Belief elicitation in the two conditions will allow us to check their invariance to the condition, that we assume here:

H.I2 [Low *vs.* High Condition on Choice]: Given the same second-order belief about *Transfer* in both conditions, the frequency of *Transfer* choices by intermediaries is the same in the Low and in the High conditions.

Third, WT in Eq. (12) is increasing in γ_{Ij}, *i.e.*, $\frac{\delta WT}{\delta \gamma_{Ij}} > 0$, and, for strictly positive second-order beliefs, in θ_{Ij}, *i.e.*, $\frac{\delta WT}{\delta \theta_{Ij}} > 0$, independently from the condition Low or High. Furthermore, Figure 3 shows that, fixing $\gamma_I < 1/2$ in the white region and moving horizontally through consecutive increases in θ_{Ij}, the intermediary switches from *Embezzle* to *Transfer* first for $\beta_{Ij} = 1$, then for $\beta_{Ij} = 2/3$, and finally for $\beta_{Ij} = 1/3$. However, even for $\theta_{Ij} \to \infty$, he will never switch from *Embezzle* to *Transfer* for $\beta_{Ij} = 0$. All this is summarized in the following hypothesis about the intermediary's **type-dependent behavior**:

H.I3 [Choice-Type Correlation]: Given the second-order belief, the frequency of *Transfer* choices increases with the altruism sensitivity; it increases with the guilt sensitivity only if the second-order belief is strictly positive. Furthermore, given a sufficiently low altruism sensitivity, the higher the guilt sensitivity, the lower the second-order belief about *Transfer* sufficient to switch from *Embezzle* to *Transfer*.

Note that we derived H.I1, H.I2, and H.I3 without specifying the treatment (Donor or Recipient), and that these hypotheses should hold in both treatments, if no treatment difference in the distribution of intermediaries' psychological types (θ_{Ij}, γ_I) were detected. Altruism being

16

a distributional preference, hence belief-independent, we expect the sensitivity to altruism not to depend on the fact that the intermediary's belief-dependent strategy relies on the donor's or the recipient's first-order beliefs. Conversely, the sensitivity to guilt might depend on the direction, *i.e.*, on whether it is elicited toward the donor or toward the recipient. Indeed, this is one of the two main research objectives of our study. However, absent previous experimental evidence on this issue, we elaborate our last hypothesis on the intermediary's **treatment-independent behavior** assuming the same distribution of guilt types across treatments:

H.I4 [Donor *vs.* Recipient treatment on Choice]: Under the same distribution of sensitivities to guilt and altruism, 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 give 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 give). The belief elicitation was incentivized. For each role, one belief was randomly selected at the end of the session and paid

€1 if accurate.[13]

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.[14] 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.

Then, 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. To facilitate decision-making, these first-order beliefs were presented in a fixed increasing order (see an example of a decision screen in Appendix C.1).[15]

---

[13]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 give 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 A11 in Appendix).

[14]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 would require too much concentration.

[15]The use of the menu method is frequent in the experimental literature on guilt aversion (Attanasi

At the end of the session, the computer program randomly selected either the Low or the High condition. Given that the donor had given a share 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

In the third part of the experiment, before subjects received any feedback on payoffs and others' decisions, 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 belief 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.[16]

The players' social norms were identified, using the Krupka and Weber (2013) procedure, for

_____

et al., 2013; Khalmetski *et al.*, 2015; Hauge, 2016; Balafoutas and Fornwagner, 2017; 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).

[16]Note that d'Adda *et al.* (2016) found no difference in responses between eliciting normative judgments *à la* Krupka and Weber (2013) before or after playing the main game.

each donors' and intermediaries' 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 A12 in Appendix).[17]

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 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 (Hypothesis H.I3) and altruism (Hypotheses H.D3, H.I3) affect behavior in the game, we elicited the subjects' social preferences by means of several psychological 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 C.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 part was included to control for poten-

---

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

tially relevant psychological traits. It 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 consisted 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 A1 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 C.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. In the second part, first the donors and then the 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 social norms and beliefs (see summary statistics and significance tests in Table A2 in Appendix). First, *Give* and *Transfer* choices are rated by the participants as significantly more socially appropriate than, respectively, *Keep* and *Embezzle* choices, in both conditions (Wilcoxon signed rank tests, W hereafter, $p < 0.001$). Second, the donors' actual (non-induced) 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 donors' and recipients' FOB on the frequency of *Transfer* choices (MW tests, $p < 0.05$ in three out of four cases).

In the following, we consider first, the donors' behavior (Section 4.1) and next, the intermediaries' behavior (Section 4.2). For each behavioral hypothesis, we check for treatment differences under the label [Donor *vs.* Recipient treatments]. Except when specified otherwise, the non-parametric tests are two-sided; an independent observation corresponds to a decision (since only one decision per participant is payoff relevant); the results from the two treatments are pooled.

4.1   Donors' Behavior

The comparison of our data to the set of 'type-belief' pairs consistent with a *R*ational donor (Eqs. (10) and (11)) shows that our model captures 92.3% of the observed behavior (see the details and the implications in terms of altruism sensitivity in Tables A4–A7 in Appendix).

Overall, 48.78% of donors chose *Give* in the Low condition and 36.56% in the High condition. Figure 4 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 A3 in Appendix). The figure illustrates our first two results on the donors' behavior.
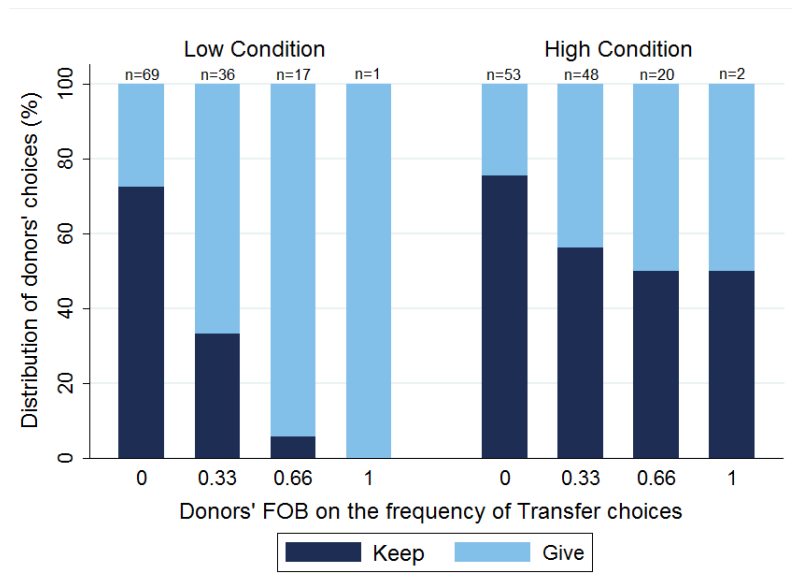
Figure 4: Distribution of the donors' choices depending on their first-order beliefs

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 in both conditions.[18]

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, $z = 3.05$, $p < 0.001$).[19]

---

[18]One may suspect that an experimenter demand effect might explain donors' giving despite the sure loss of material payoff entailed by the *Give* choice. However, the detected positive correlation between donors' *Give* choices and their first-order belief of intermediary's *Transfer* choices makes us confident that an experimenter demand effect is not the main driver of donors' behavior.

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

23

[Donor *vs.* Recipient treatments]: The correlation between the donors' FOB on the frequency of *Transfer* choices and their decision to *Give* is not significantly different across treatments (Donor treatment: $r_s = 0.44$, $p < 0.001$; Recipient treatment: $r_s = 0.24$, $p < 0.001$; 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, $p = 0.057$).[20]

[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, $p = 0.252$).

Result D3 [Choice-Type Correlation]: Controlling for the donors' FOB about *Transfer*, the frequency of *Give* choices tends to increase with the donor's sensitivity to altruism. This holds in both conditions.

Support for Result D3: We use the Self-Reported Altruism score (Rushton *et al.*, 1981)

Although belief elicitation was incentivized, it is possible that donors who planned to *Keep* may have underestimated their FOB about *Transfer* choices to justify their selfish choice. To further test H.D1, 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, $r_s = -0.20$, $p < 0.001$).

[20]Two donors had a FOB of 0.66 in both conditions and no donor had a FOB of 1 in both conditions.

as a proxy for our altruism sensitivity parameter. For a given FOB of 0.33, there is a marginally significant positive correlation between the donors' *Give* choices and their altruism score (S correlation, $r_s = 0.42$, $p = 0.081$). However, the correlation is not significant when the given FOB is 0 (S correlation, $r_s = 0.15$, $p = 0.313$). As for Result D2, we cannot test our hypothesis with the other two FOB.[21]

[Donor *vs.* Recipient treatments]: The correlation between the donors' *Give* choices and their altruism score is not different across treatments (Z test, FOB(0): $z = -0.93$, $p = 0.176$; FOB(0.33): $z = 0.61$, $p = 0.270$).

4.2 Intermediaries' Behavior

The comparison of our data to the set of 'type-belief' pairs consistent with a Rational intermediary (Eq. (13)) shows that our model captures 82.93% of the observed behavior: (i) 46.75% of the intermediaries always chose *Embezzle*, (ii) 24.39% switched from *Embezzle* to *Transfer* as the induced SOB increases, *i.e.*, exhibiting guilt aversion, and (iii) 11.79% always chose *Transfer*, *i.e.*, exhibiting altruistic preferences prevailing over guilt aversion.[22]

Focusing on behavior consistent with our theoretical predictions, Figure 5 presents the distribution of the switching SOB observed in the two treatments and the implications in terms of predicted altruism sensitivity and guilt sensitivity (see Figure 3). The distributions of switching SOB do not differ significantly across treatments (Kruskal-Wallis test, $p > 0.10$).

---

[21]Two donors had a FOB of 0.66 in both conditions and no donor had a FOB of 1 in both conditions.

[22]Figure 3 shows that intermediaries who always choose *Transfer* for any second-order belief have a sensitivity to altruism $\gamma_I > 1/2$, but they can also have a sensitivity to guilt $\theta_{Ij} > 0$. Since we do not know how many of them have $\theta_{Ij} > 0$, the fraction of intermediaries exhibiting guilt aversion might be underestimated in our sample of participants.
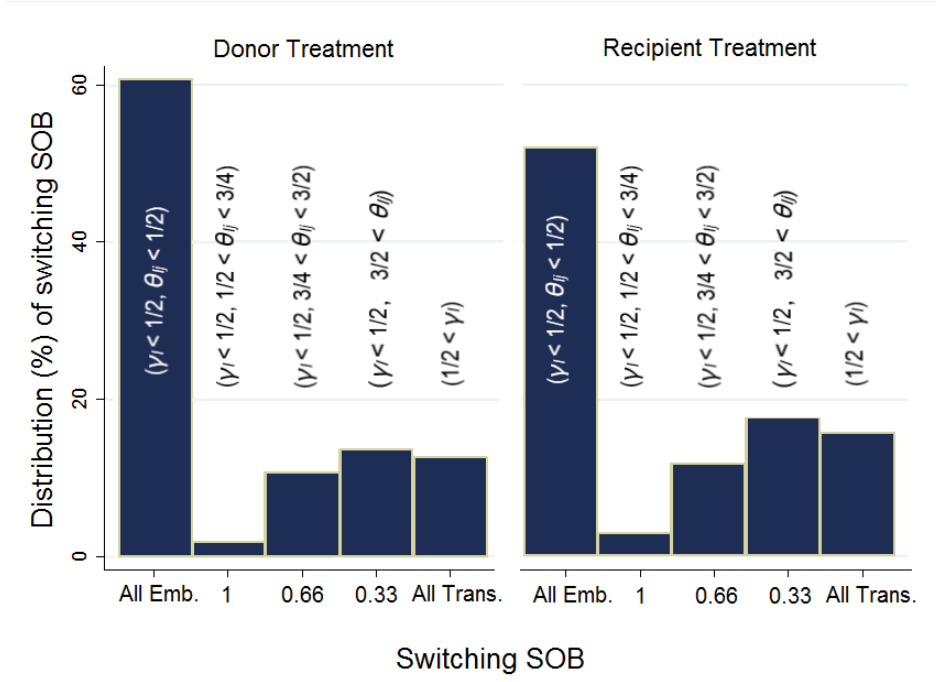
Figure 5: Distribution of the intermediaries' switching second-order beliefs

Notes: Subjects who did not behave consistently with our theoretical predictions are excluded from this figure. The figure reads as follows. In the Recipient Treatment, 3% of the intermediaries have a switching SOB of 1, *i.e.*, they chose *Embezzle* for an induced SOB in $\{0; 0.33; 0.66\}$ and *Transfer* for an induced SOB of 1. This behavior is consistent with guilt aversion prevailing over altruism only if $\gamma_I < \frac{1}{2}$ and $\frac{1}{2} < \theta_{Ij} < \frac{3}{4}$ (see Figure 3).

The remaining intermediaries behave as follows: 11.79% of intermediaries switched multiple times between transferring and embezzling, and 5.28% exhibit an inverse switching pattern from transferring to embezzling.

Figure 6 displays, for each condition, the proportion of intermediaries who chose either *Transfer* or *Embezzle*, depending on their induced SOB (see also Table A8 in Appendix). The figure illustrates our first two results on the intermediaries' behavior.
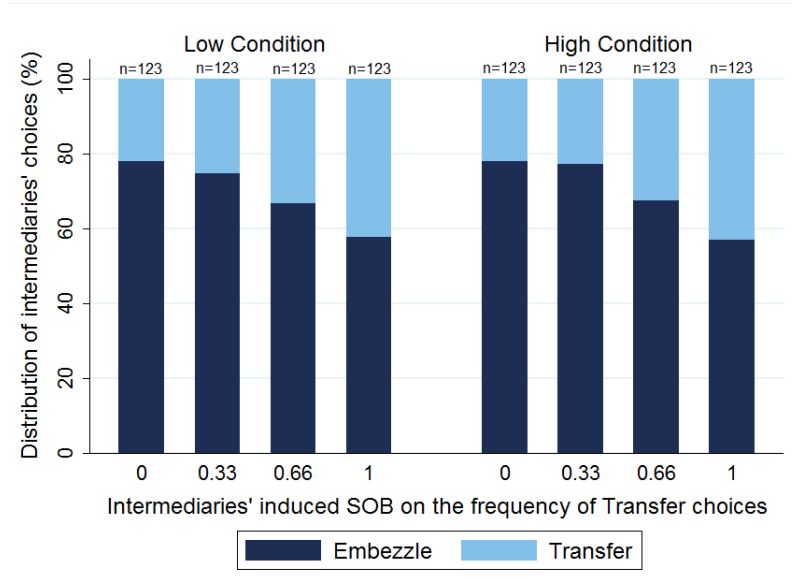
Figure 6: Distribution of the intermediaries' choices depending on their induced second-order beliefs

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

Support for Result I1: There is a significant positive correlation between the intermediaries' induced SOB about *Transfer* and their *Transfer* choices (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 a hypothetical bias, as they are sure that their choices will not be payoff-relevant, rendering the hypothetical decision to embezzle less psychologically costly.

So far, we have conducted the analysis by examining the induced SOB based on the menu method of Khalmetski *et al.* (2015). If, instead, we use the stated SOB (second-order beliefs reported directly by the subjects in the third part of the experiment), we find that the correlation between the intermediaries' *Transfer* choices and their stated SOB 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 A9 in Appendix) and with random effects and individuals controls (Table A10 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 Tables A9 and A10 in Appendix).

Result I2 [High *vs.* Low condition on Choice]: Controlling for the intermediaries' induced SOB about *Transfer*, the frequency of *Transfer* choices is the same in both conditions.

Support for Result I2: We use MN tests to consider each intermediary as an independent observation. For a given induced SOB, the frequency of *Transfer* choices does not significantly differ across conditions (smallest $p = 0.438$) (see also Table A10).

[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: $\chi^2 = 4.50$, $p = 0.033$) (see also Table A8 in Appendix).

Result I3 [Choice-Type Correlation]: The frequency of *Transfer* choices increases (i) for a given second-order belief, with the altruism sensitivity, and (ii) for a given second-order belief, with the guilt sensitivity. Furthermore, the higher the guilt sensitivity, the lower the second-order belief about Transfer sufficient to switch from *Embezzle* to *Transfer*. This holds in both conditions.

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

---

[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.*, 2017).

parameter in our model and the Self-Reported Altruism score (Rushton *et al.*, 1981) as a proxy for the altruism-sensitivity parameter. Table 1 presents (i) the correlation between the *Transfer* choices, holding the induced SOB constant, and the Guilt-NBE score, (ii) the correlation between the switching SOB and the Guilt-NBE score, as well as (iii) the correlation between the *Transfer* choices, holding the induced SOB constant, and the Altruism score. The switching SOB corresponds to the minimum induced SOB sufficient to choose *Transfer* rather than *Embezzle*.[24]

	Donor treatment			Recipient treatment		
	Guilt	Guilt x Hypoth.	Z-stat	Guilt	Guilt x Hypoth.	Z-stat
Transfer $\mid$ SOB=0	0.10	0.27**	-0.95	0.17	0.23*	-0.34
Transfer $\mid$ SOB=0.33	0.19	0.39***	-1.18	0.43***	0.46***	-0.20
Transfer $\mid$ SOB=0.66	0.15	0.47***	-1.93**	0.19*	0.43***	-1.44*
Transfer $\mid$ SOB=1	0.11	0.49***	-2.29**	0.23*	0.52***	-1.84**
Switching SOB	-0.14	-0.49***	2.07***	-0.10	-0.34***	1.33***
	Altruism	Altruism x Hypoth.	Z-stat	Altruism	Altruism x Hypoth.	Z-stat
Transfer $\mid$ SOB=0	-0.01	0.28**	-1.60*	0.06	0.14	-0.44
Transfer $\mid$ SOB=0.33	-0.01	0.35***	-2.02**	0.23***	0.27***	-0.23
Transfer $\mid$ SOB=0.66	0.02	0.46***	-2.57***	0.03	0.25**	-1.21
Transfer $\mid$ SOB=1	-0.08	0.44***	-2.97***	0.02	0.39***	-2.11**

Notes: This table presents the coefficients of Spearman correlations between row and column variables. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Rows: "Transfer $\mid$ SOB $= \beta_{Ij}$" represents the total of *Transfer* choices in both conditions given that the induced second order belief was β_{Ij}. "Switching SOB" represents the switching SOB of intermediaries who either always *Embezzle* or are guilt-averse.
Columns: Guilt stands for Guilt-NBE score (GASP questionnaire). Altruism stands for self-reported altruism score (Rushton *et al.*, 1981). Hypoth. stands for a dummy variable that takes value 0 if the intermediary believes that no donor will choose *Give* in either condition, and 1 otherwise. Z-stat stands for the differences between columns measured by Fisher r-to-Z transformations (one-tailed).

Table 1: Correlation between the intermediaries' decisions and their Guilt-NBE and Altruism scores

The Guilt-NBE score in itself is only marginally significantly correlated with *Transfer* choices.

---

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

However, the strength of the correlation improves if we interact this score with a dummy variable that takes value 0 if the intermediary believes that no donor will choose *Give* in either condition, and 1 otherwise (see Table 1). The same remark holds for Altruism score, although the improvement is significant only in the Donor Treatment (see Table 1). This suggests that the Guilt-NBE score and the Self-Reported Altruism score are relevant proxies for, respectively, the guilt-sensitivity and altruism-sensitivity parameters only when intermediaries believe their decision will be implemented with a non-null probability.

[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 I1–I3 above, see (the absence of) treatment difference under the label [Donor *vs.* Recipient treatments].

4.3 A Structural Estimate of Guilt Sensitivity

Following Bellemare *et al.* (2011), we define a structural econometric model to estimate the intermediaries' average guilt-sensitivity parameter, θ_{Ij}, toward the donor ($j = D$, in the Donor treatment) and the recipient ($j = R$, in the Recipient treatment).[25] Given the treatment, for each α_{jI} and each condition (eight cases per intermediary), intermediaries choose s_I (*Transfer* or *Embezzle*) to maximize their utility after *Give*, as defined by Equation (14) (Random Utility

---

[25]Recall that, in our theoretical model, we also introduce a parameter which represents the altruism sensitivity, γ_I (see Eq. (7)). However, the second component of the intermediary's feeling of altruism A_{IR}, *i.e.*, the recipient's received amount $r(s_I)$, is colinear with the intermediary's material payoff (by construction of the Mini-Games): $r(s_I) = 2 \cdot [25 - (M_I(s_I) - 80)]$ in both conditions. Therefore, we cannot estimate the three coefficients (γ_I, θ_{Ij}, and the coefficient corresponding to M_I) of our theoretical utility function, since our data set contains only two variables that are not colinear (M_I and B_{Ij}). A solution would be to force the coefficient corresponding to M_I to 1, however we could not estimate the noise parameter of our random utility model (Eq. (14)). Therefore, we renounce to estimate γ_I.

Model). In this equation, λ is the noise parameter that we estimate, and U_I is defined following our modeling of guilt aversion toward the donor or the recipient (Eq. 6): $U_I(\theta_{Ij}, s_I | s_D = G, \alpha_{jI}) = 1 \cdot M_I(G, s_I) - \theta_{Ij} \cdot \max\{0, \mathbb{E}_j[r(s_I)|s_D = G] - r(s_I)\}$ in the Low and High condition, for $\alpha_{jI} \in \{0, 1/3, 2/3, 1\}$, and $j \in \{D, R\}$:

$$V_I(\theta_{Ij}, \lambda, s_I) \quad = \quad U_I(\theta_{Ij}, s_I) + \lambda \cdot \epsilon_I(s_I) \tag{14}$$

We used a conditional Logit model to estimate θ_{Ij}, the coefficient corresponding to the guilt sensitivity parameter, and λ, the noise parameter, while fixing to 1 the coefficient corresponding to the intermediary's own material payoff. Table 2 reports the results of these estimates.

	All treatments	Donor treatment	Recipient treatment	Subjects with Hypothetical Bias Excluded
θ_{Ij}	-0.37***	-0.34***	-0.41***	-0.61***
	(0.03)	(0.04)	(0.04)	(0.04)
λ	6.48***	5.80***	7.32***	8.30
	(0.42)	(0.51)	(0.74)	(0.79)
N	123	62	61	83

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

Table 2: Structural estimates of the guilt-sensitivity parameter

The results reported in Table 2 show that the average intermediary is willing to pay 0.37 ECU to avoid letting down another player by 1 ECU (difference between expectations and actual outcome). When we exclude intermediaries who believed that no donor chose to *Give* (those intermediaries who are potentially subject to a hypothetical bias), the estimated guilt-sensitivity parameter increases up to 0.61.

[Donor *vs.* Recipient treatments]: Although intermediaries seem to be slightly more sensitive to guilt toward the recipient than toward the donor (+20%), the difference is not significant (Z test, $z = -1.09$, $p = 0.13$; see Paternoster *et al.*, 1998).

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 (ii) whether the impact of others' expectations on behavior differs if others are the donors or the potential recipients of the donation. Extending Battigalli and Dufwenberg (2007) model to capture guilt aversion toward the donor and documenting its existence and features in a laboratory experiment are our two original contributions. Indeed, we have modeled 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. 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. We are the first to address these three questions in a single paper.

We find that (i) on average, about 25 % of the 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 guilt; (ii) on average, an intermediary is willing to pay 0.37 ECU not to let down another player by 1 ECU, and the intensity of the structurally estimated guilt-sensitivity parameter is not significantly different when the intermediary is confronted with the recipient's expectations (0.41) compared to the donor's expectations (0.34). Thus, guilt aversion has the same effect on intermediaries, regardless of whether the intermediary considers a person that may be financially harmed by his decision or a person that he may betray but without any monetary consequences.

Our results contribute to 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 B1 in Appendix). Our structural estimates of guilt sensitivity are in the same range of values as those obtained by Bellemare *et al.* (2011, 2018) through structural estimations (see

Table B2 in Appendix). 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 believe that their decision will be implemented with a non-null probability that they are not playing hypothetically. This finding contributes to the small literature trying to identify the link between survey-based measures of guilt and experimental decisions (see Table B3 in Appendix). Overall, this calls for more research on the nature of the emotions embedded in the Battigalli and Dufwenberg (2007) guilt-aversion model.

These findings highlight that psychological game theory can contribute usefully to the renewal of the analysis of dishonesty by a better understanding of the moral costs of unethical behavior. 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 complex design, though. By enlarging the perspective to a dynamic setting, we could also contribute to explain the emergence of a vicious circle of corrupt norms. If donors or recipients expect a high level of embezzlement in a group, intermediaries can embezzle without feeling guilty, which in turn increases the expectations of embezzlement.

If the results on intermediaries' guilt aversion in the lab hold in the field, anti-corruption policies could publicize the high expectations of donors and recipients to the intermediaries. Public campaigns of information (Reinikka and Svensson, 2011) or framing manipulations (Ockenfels and Werner, 2014) usually focus on the potential recipients' expectations. Policies should also consider the sensitiveness of intermediaries to the donors' expectations (see also the literature on trust-responsiveness, *e.g.*, Guerra and Zizzo, 2004; Bacharach *et al.*, 2007). But of course, identifying guilt aversion in the lab does not prove that it exists to the same extent in the field. In the field there may be an asymmetry in guilt aversion because the hierarchy of status or power adds to the inequality of payoffs that we introduced in our experiment (for example, donors are sometimes a corrupt and exploitative government dealing with other people's money; thus, guilt

toward the recipients may be much stronger than toward the donor). Note that previous studies on bribery found no difference in behavior in the field and in the lab (Armantier and Boly, 2013) and that dishonesty in the lab correlates with dishonesty of the same individuals in the field (Dai *et al.*, 2018). Future research should usefully test the qualitative and quantitative external validity of our results. A major challenge, though, will be to measure beliefs in the field.

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 HEX-ACO 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., 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., Dufwenberg, M. (2018a). Incorporating belief-dependent motivation in games. Typescript, Bocconi University.

Battigalli, P., Corrao, R., Sanna, F. (2018b). 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.

Beekman, G., Bulte, E., Nillesen, E. (2014). Corruption, investments and contributions to public goods: Experimental evidence from rural Liberia. *Journal of Public Economics 115*, 37-47.

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*, 21, 316-336.

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.

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.

Dufwenberg, M., Dufwenberg, M.A. (2018). Lies in disguise – A theoretical analysis of cheating. *Journal of Economic Theory 175*, 248-264.

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.

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

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review 95 (1)*, 384–394.

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

Kajackaite, A., Gneezy, U. (2017). Incentives and cheating. *Games and Economic Behavior 102*, 433-444.

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.

Köbis, N.C., van Prooijen, J.-W., Righetti, F., Van Lange, P.A.M. (2016). Prospection in individual and interpersonal corruption dilemmas. *Review of General Psychology 20(1)*, 71-85.

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 & 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. *Journal of Economic Behavior Organization.* Forthcoming.

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.

A   Additional Tables

Session (#)	Participants (n)	Age (mean)	Women (%)	Previous Exp. (mean)	Economics Stud. (%)
Donor Treatment					
5	18	21.27	50.00	1.44	66.67
6	18	21.00	72.22	0.22	33.33
7	15	24.20	53.33	1.53	33.33
8	21	22.00	90.48	0.80	28.57
9	21	21.28	42.86	1.80	71.43
12	24	23.25	50.00	1.33	45.83
13	18	20.66	55.56	1.50	61.11
15	21	21.19	47.82	1.00	38.10
16	12	21.00	50.00	2.50	50.00
18	15	22.40	46.67	1.86	33.33
Sub-total	183	21.83	56.28	1.34	46.45
Recipient Treatment					
1	18	21.77	22.22	1.16	77.78
2	21	19.76	61.90	0.90	57.14
3	15	20.93	40.00	0.26	80.00
4	21	20.85	57.14	1.14	52.38
10	18	22.88	72.22	1.16	61.11
11	27	22.30	50.00	1.96	62.96
14	24	21.50	37.50	2.20	54.17
17	27	24.59	40.74	2.70	55.56
19	15	21.00	66.67	2.46	46.67
Sub-total	186	21.87	49.18	1.63	60.22
Treatment Difference		No[1]	No[2] No[1]	Yes[2]***	
Total	369	21.85	52.72	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 A1: 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 Give [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 A2: 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 A3: Donors' *Give* choices for a given FOB on the frequency of *Transfer* choices

(*Keep, Keep*)	FOB = 0		FOB = 0.33		FOB = 0.66		FOB = 1	
	Prediction	n	Prediction	n	Prediction	n	Prediction	n
FOB = 0	$\gamma_D < \frac{5}{4}$	34	$\gamma_D < \frac{15}{14}$	10	$\gamma_D < \frac{15}{22}$	1	$\gamma_D < \frac{1}{2}$	0
FOB = 0.33	$\gamma_D < \frac{5}{6}$	4	$\gamma_D < \frac{5}{6}$	4	$\gamma_D < \frac{15}{22}$	3	$\gamma_D < \frac{1}{2}$	0
FOB = 0.66	$\gamma_D < \frac{5}{8}$	0	$\gamma_D < \frac{5}{8}$	1	$\gamma_D < \frac{5}{8}$	0	$\gamma_D < \frac{1}{2}$	0
FOB = 1	$\gamma_D < \frac{1}{2}$	0	$\gamma_D < \frac{1}{2}$	0	$\gamma_D < \frac{1}{2}$	0	$\gamma_D < \frac{1}{2}$	0

Example: 10 donors chose to *Keep* in the Low condition and to *Keep* in the High condition while having a FOB of 0 about *Transfer* choices in the Low condition and a FOB of 0.33 about *Transfer* choices in the High condition. This behavior is consistent with our theoretical predictions only if $\gamma_D < \frac{15}{14}$.

Table A4: Matching the donors' behavior to our predictions - (*Keep, Keep*)

(*Give, Keep*)	FOB = 0		FOB = 0.33		FOB = 0.66		FOB = 1	
	Prediction	n	Prediction	n	Prediction	n	Prediction	n
FOB = 0	$\frac{5}{4} < \gamma_D < \frac{5}{2}$	0	x	1	x	0	x	0
FOB = 0.33	$\frac{5}{6} < \gamma_D < \frac{5}{2}$	2	$\frac{5}{6} < \gamma_D < \frac{15}{14}$	7	x	5	x	0
FOB = 0.66	$\frac{5}{8} < \gamma_D < \frac{3}{2}$	0	$\frac{5}{8} < \gamma_D < \frac{15}{14}$	4	$\frac{5}{8} < \gamma_D < \frac{15}{22}$	1	x	1
FOB = 1	$\frac{1}{2} < \gamma_D < \frac{5}{2}$	0	$\frac{1}{2} < \gamma_D < \frac{15}{22}$	0	$\frac{1}{2} < \gamma_D < \frac{15}{14}$	0	x	0

Example: see Table A4

Table A5: Matching the donors' behavior to our predictions - (*Give, Keep*)

(*Give, Give*)	FOB = 0		FOB = 0.33		FOB = 0.66		FOB = 1	
	Prediction	n	Prediction	n	Prediction	n	Prediction	n
FOB = 0	$\frac{5}{2} < \gamma_D$	11	$\frac{5}{4} < \gamma_D$	4	$\frac{5}{4} < \gamma_D$	3	$\frac{5}{4} < \gamma_D$	0
FOB = 0.33	$\frac{5}{2} < \gamma_D$	0	$\frac{15}{14} < \gamma_D$	6	$\frac{5}{6} < \gamma_D$	4	$\frac{5}{6} < \gamma_D$	0
FOB = 0.66	$\frac{5}{2} < \gamma_D$	1	$\frac{15}{14} < \gamma_D$	7	$\frac{15}{22} < \gamma_D$	1	$\frac{5}{8} < \gamma_D$	1
FOB = 1	$\frac{5}{2} < \gamma_D$	0	$\frac{15}{14} < \gamma_D$	0	$\frac{15}{22} < \gamma_D$	1	$\frac{1}{2} < \gamma_D$	0

Example: see Table A4

Table A6: Matching the donors' behavior to our predictions - (*Give, Give*)

(*Keep*, *Give*)	FOB $= 0$		FOB $= 0.33$		FOB $= 0.66$		FOB $= 1$	
	Prediction	n	Prediction	n	Prediction	n	Prediction	n
FOB $= 0$	x	1	$\frac{15}{14} < \gamma_D < \frac{5}{4}$	3	$\frac{15}{22} < \gamma_D < \frac{5}{4}$	1	$\frac{1}{2} < \gamma_D < \frac{5}{4}$	0
FOB $= 0.33$	x	0	x	1	$\frac{15}{22} < \gamma_D < \frac{5}{6}$	0	$\frac{1}{2} < \gamma_D < \frac{5}{6}$	0
FOB $= 0.66$	x	0	x	0	x	0	$\frac{1}{2} < \gamma_D < \frac{5}{8}$	0
FOB $= 1$	x	0	x	0	x	0	x	0

Example: see Table A4

Table A7: Matching the donors' behavior to our predictions - (*Keep*, *Give*)

		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 A8: Intermediaries' *Transfer* choices for a given induced SOB

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

Notes: In the first four columns, the dependent variable is the decision to *Transfer* made for a given induced SOB (using the menu method). In the last column, the dependent variable is the decision to *Transfer* made when the induced SOB corresponded to the stated SOB (the SOB reported directly by the intermediaries in the third part of the experiment). Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A9: Regression on the decision to Transfer (Logit model with fixed effects)

	All treatments	Donor treatment	Recipient treatment	Hypothetical Bias Excluded	All treatments
Induced SOB	0.75***	0.74***	0.76***	0.90***	
	(0.10)	(0.13)	(0.15)	(0.11)	
Low Condition	0.08	0.27	-0.13	0.13	-0.04
	(0.20)	(0.28)	(0.30)	(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 A10: 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 A11: 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 A12: Kruskal-Wallis tests of the difference in social norms distributions across roles

B   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 B1: 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	0.34	61
			Toward Recipient	0.41	62

Table B2: Previous estimations of the guilt-sensitivity parameter

		Correlation between ...			
Study	Game	Trait	Behavior	$p < 0.1$	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 B3: Previous correlation of personality traits and behavioral outcomes

44

C Online Appendices

C.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

<div align="center">ROLE OF THE DONOR</div>

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.

<div align="center">ROLE OF THE INTERMEDIARY</div>

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 some questions.

- Stage 2: The donor makes his/her decisions.

- Stage 3: The intermediary makes his/her decisions.

- Stage 4: All the participants answer 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 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 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 two questions about the intermediaries' decisions (in situation A and in situation B);

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

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

How do the answers affect your earnings?

At the end of the session, for each role, one of the questions to which you have answered will be randomly selected. If your answer to that question corresponds to what truly happened, you will earn 1.

Example: Suppose you are a recipient and the question randomly selected is "In situation B, among 3 donors randomly selected in today's session, in your opinion, how many of these donors

48

will choose to give 25 ECU toward the recipient?". The program randomly select 3 donors among the participants to todays' session. If in situation B, x donor(s) among the 3 randomly selected ones, has/have given 25 ECU toward the recipient, then you answer is correct if you answered "x".

STAGE 2

In this stage, the donors make their decisions.

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

<u>If you are a donor</u>, 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.

<center>***</center>

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

<center>*[The next set of instructions was distributed after stages 1 and 2.]*</center>

<center>**STAGE 3**</center>

**In this stage, the intermediaries make their decisions.**

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

<u>If you are an intermediary</u>, 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.

Situation B

If your donor thinks that **1 out of 3** intermediary randomly selected today will transfer 25 ECU .

How many ECU do you wish to transfer?
- ○ 25 ECU
- ○ 5 ECU

Continuer

Figure 7: Screenshot for the *Donor Treatment*

<center>50</center>
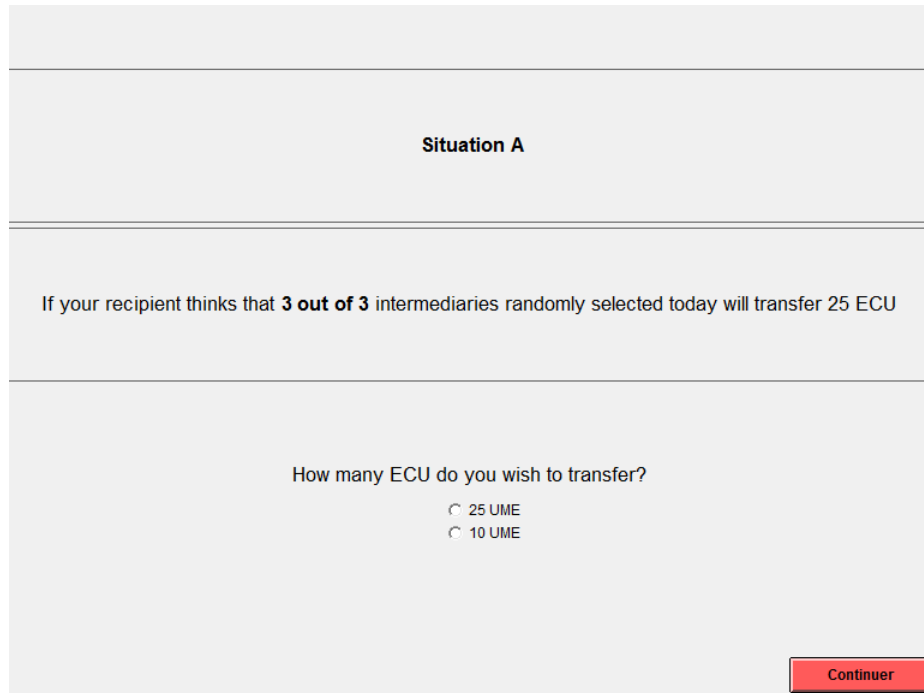
Figure 8: Screenshot for the *Recipient Treatment*

- **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:

- <u>Regardless of the situation</u>, if the intermediary transfers 25 ECU to the recipient, the intermediary's payoff is 80 ECU and the recipient's payoff is 60 ECU.

- <u>If situation A is randomly selected</u> 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.

- <u>If situation B is randomly selected</u> 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 a 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 <u>question of type 1</u>:

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 your questions in private.

C.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 B4: 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 B5: 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) <u>Professional Activity</u> - *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. <u>Number of previous experiments</u>
In how many GATE-LAB experimental sessions have you participated already?

8. <u>Personal Login</u>
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".