GUILT AVERSION IN (NEW) GAMES: THE ROLE OF VULNERABILITY

Documents de travail GREDEG
GREDEG Working Papers Series

Giuseppe Attanasi
Claire Rimbaud
Marie-Claire Villeval

GREDEG WP No. 2020-15
https://ideas.repec.org/s/gre/wpaper.html

Guilt Aversion in (New) Games: the Role of Vulnerability

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

GREDEG Working Paper No. 2020–15

Abstract:
From the literature we know that contextual factors modulate guilt aversion, such as pre-play communication and social closeness. In this study, we investigate whether a particular feature of the game itself — the vulnerability of the co-players — affects a player's guilt aversion. We deem that a co-player is (i) ex-post vulnerable when her final payoff depends on the decision-maker's actions, and (ii) ex-ante vulnerable when the use of her initial endowment depends on the decision-maker's actions. In a laboratory experiment, we introduce four (new) three-player trust games played within-subjects, varying whether the trustees can condition their decision on the belief of another player who is ex-post vulnerable and/or ex-ante vulnerable. We put forward a portable model of lexicographic altruism and role-dependent guilt, where the trustee can only be altruistic toward the most disadvantaged player and can feel guilty simply because of his role in the game. We find that trustees' guilt aversion is insensitive to the opponents' vulnerability and to the role of the vulnerable player.

JEL codes: C72, C91, D91

Keywords: Guilt aversion, vulnerability, psychological game theory, experiment

[a] Université Côte d'Azur, GREDEG UMR 7321, 250 Rue Albert Einstein, F-06560, Valbonne, France. E-mail: giuseppe.attanasi@unice.fr
[b] Univ Lyon, University Lyon 2, GATE UMR 5824, 93 Chemin des Mouilles, F-69130, Ecully, France. E-mail: rimbaud@gate.cnrs.fr
[c] Univ Lyon, CNRS, GATE UMR 5824, 93 Chemin des Mouilles, F-69130, Ecully, France. IZA, Bonn, Germany. E-mail: villeval@gate.cnrs.fr

Acknowledgements: We are grateful to G. Andrighetto, M. Dufwenberg, A. Guido, E. Manzoni, S. Papa and L. Tummolini for very valuable feedback. We thank also participants at the 3rd Workshop on Psychological Game Theory (Soleto), the Workshop on Ethics and Emotions (Paris), the 1st CoCoLab Workshop (Nice), the EDSEG Doctoral Workshop (Lyon), the ASFEE Conference (Toulouse), the ESA World Conference (Vancouver), and the Laboratory of Agent-Based Social Simulation ISTC-CNR (Rome) for very useful comments. The research leading to these results has received funding from the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-007) operated by the French National Research Agency (ANR). This research has also benefited from the support of the French Agence Nationale de la Recherche (ANR) under grant ANR-18-CE26-0018-01 (project GRICRIS).

1 Introduction

”If people feel guilt for hurting their partners [...] and for failing to live up to their expectations, they will alter their behavior (to avoid guilt) in ways that seem likely to maintain and strengthen the relationship.” (Baumeister et al., 1994, p. 247). Based on these psychological insights, economists have modelled how guilt can influence actions in the framework of psychological game theory (see Battigalli and Dufwenberg, 2019, for a survey).[1] Building on the seminal work of Geanakoplos et al. (1989), Battigalli and Dufwenberg (2007) have introduced the standard definition of guilt aversion as a belief-dependent motivation, which states that an agent suffers a psychological cost, *i.e.*, feels guilty, if he lets down others' expectations. Guilt aversion has been shown to promote cooperation in trust games (*e.g.*, Charness and Dufwenberg, 2006; Reuben et al., 2009; Bellemare et al., 2017; see for a review Cartwright, 2018) and sender-receiver games (*e.g.*, Battigalli et al., 2013); and to motivate prosocial behavior in dictator games (*e.g.*, Ockenfels and Werner, 2014; Hauge, 2016; Bellemare et al., 2018). Moreover, it has been applied to a variety of situations from corruption (Balafoutas, 2011; Attanasi et al., 2019b) to drug use in sports (Cartwright, 2019).

Several papers have investigated what factors could modulate guilt aversion. For instance, the possibility to communicate among players greatly facilitates the expression of guilt aversion. It has been shown in the milestone paper of Charness and Dufwenberg (2006) and replicated in many experimental papers since (*e.g.*, Attanasi et al., 2013; Bracht and Regner, 2013, Kawagoe and Narita, 2014; Balafoutas and Fornwagner, 2017; Attanasi et al., 2019a). Further studies aimed to understand why communication encourages guilt-averse behaviors. First, the role of promises exchanged during the communication has received a lot of attention (*e.g.*, Vanberg, 2008; Beck et al., 2013; Ederer and Stremitzer, 2017; Chen and Houser, 2018; Di Bartolomeo et al., 2018; Di Bartolomeo et al., 2019), with only some of these studies reporting clear-cut evidence in favor of guilt aversion. Second, it has been hypothesized that communication could facilitate the expression of guilt aversion through an increased social closeness between players. Supporting this intuition, in the absence of communication, Morell (2014) has shown that the proportion of guilt-averse players was higher among socially-close players (same social identity).

It has also been shown that the nature of expectations greatly influences the emergence of guilt aversion. First, “reasonable” expectations are more likely to be taken into account by guilt-averse players. Balafoutas (2011), Khalmetski (2016) and Danilov et al. (2019) all reported an inverse-U shaped relationship between second-order beliefs and actions: dictators are less prosocial when they deem that recipients expect to receive too few or too much. Pelligra et al. (2016) partially confirmed this result in a

[1]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.

trust game including an experimenter's requests to the trustees: low requests led to lower return but high requests tended to be ignored. Second, in a trust mini-game, Andrighetto et al. (2015) showed that the emergence of guilt aversion is facilitated by the perceived legitimacy of others' normative expectations. Third, in a sender–receiver game, Peeters et al. (2015) found that guilt aversion is a relevant channel to explain truth-telling in social dilemmas where the expectation of truth-telling has a strong normative appeal, but not in antagonistic games where the normative appeal is weaker.

Finally, based on sender-receiver and dictator games, Khalmetski (2016) and Bellemare et al. (2018) have shown that the dictator's guilt aversion is influenced by the size of the game payoffs. In fact, both find that dictators' average guilt sensitivity is significantly decreasing with the level of stakes.

However, beyond contextual factors and level of stakes, no study has investigated whether the structure of the game itself can modulate the emergence of guilt aversion. In particular, we contend that the game structure can determine the vulnerability of co-players to the decision-maker's choices. Whether this vulnerability can impact the decision-maker's guilt aversion still remains an open question. Recent studies by Cox et al. (2016) and Engler et al. (2018) showed that the second-mover's trustworthiness increases in the first-mover's vulnerability, incidentally underlining the relevance of the vulnerability of co-players in decision making.[2] Nevertheless, neither of these studies assessed the impact of vulnerability on guilt aversion, since their modelling of conditional other-regarding preferences rested on the premises of a revealed-intention approach rather than on belief-dependent preferences.

The current study was instead designed to contrast different conditions of co-players' vulnerability to assess its influence on the decision-maker's guilt aversion. By looking at whether the structure of the game itself can facilitate or attenuate the expression of guilt aversion, we contribute to the experimental literature on its determinants.

The introduction of four novel three-players games, representing a broad set of variations of a Trust Minigame with a passive player, allows us to study two types of vulnerability: ex-post vulnerability and ex-ante vulnerability. We define a player as *ex-post vulnerable* if her final payoff depends on the actions of the decision-maker. This type of vulnerability is the most intuitive and corresponds to the original modelling of guilt aversion by Battigalli and Dufwenberg (2007). We define a player as *ex-ante vulnerable* if the use of (part of) her initial endowment depends on the decision-maker's actions. These two types

[2] Cox et al. (2016) considered that the first-mover is vulnerable if she made a choice such that the maximum payoff she can obtain—assuming that the second-mover is selfish—is lower than the maximum payoff she could have obtained otherwise—again assuming a selfish second-mover. Engler et al. (2018) defined three degrees of vulnerability of the first-mover in a trust game: the first mover is either (i) not vulnerable, if she made a choice such that the minimum payoff she can obtain by going *In* is higher than the payoff she could have obtained by exiting the game; (ii) vulnerable, if she made a choice such that the two payoffs she can obtain by going *In* are respectively lower and higher than the payoff she could have obtained by exiting; (iii) very vulnerable, if she made a choice such that the maximum payoff she can obtain by going *In* is lower than the payoff she could have obtained by exiting.

of vulnerability are at play in embezzlement games such as the one studied by Attanasi et al. (2019b). Think about the following two-stage game: a rich donor (trustor) wants to give part of her revenue to a poor recipient in a foreign country (passive player); an intermediary (trustee) is supposed to transfer the donation to the recipient but he can embezzle part of the donation for himself. Here, the donor is ex-ante vulnerable since the use of her donation is in the intermediary's hands, and the recipient is ex-post vulnerable since her final wealth depends on the intermediary's choice.

Besides a Donation game close to the embezzlement game in Attanasi et al. (2019b), in the current study we let experimental participants play three other two-stage games with two active and one passive player (C): the Investment game, the Reversed-Investment game, and the Exploitation game. In each of these four three-players games, the second mover (player B) can be entrusted by the first mover (player A) with a sum of money coming from the endowment of another player (A or C, depending on the game); then, he can redistribute this sum of money between himself and another player (A or C, depending on the game). The four games are highly comparable, since they share several similar strategic features: same set of strategies of each of the two active players, same initial endowment of each of the three players, same payoff for player B given the game terminal node. As for game manipulations, the ex-ante vulnerable player and the ex-post vulnerable player can be the same person or not depending on the game. As for treatment manipulation, B's belief-dependent behavior is elicited by relying on A's first-order beliefs in one treatment and on C's first-order beliefs in another treatment.

The within-game and between-treatment comparisons of our design allow us to address several research questions about the (in)dependence of second-mover's guilt aversion on another player's vulnerability. First, Cox et al. (2016) and Engler et al. (2018) found the co-players' vulnerability to have an impact on the second-mover's behavior. In line with their results, one may contend that the co-players' vulnerability may boost the expression of the second-mover's guilt aversion. In our games, we argue that player B can feel guilty both towards the person whose endowment was the source of the money entrusted to him (ex-ante vulnerability) and toward the person who may receive the money after his redistribution choice (ex-post vulnerability). The relevance of distinguishing two types of vulnerability is an empirical question that cannot be answered *a priori*. Indeed, a single study (Attanasi et al., 2019b) observed evidence of guilt aversion when a co-player was ex-ante vulnerable, and only in the Donation game. We seek at dissociating this result from the specific strategic interaction, thereby checking whether it also holds in other two-stage games which are strategically equivalent for a selfish second mover.

Second, the relevance of the opponent's vulnerability on the second-mover's choice detected in Cox et al. (2016) and Engler et al. (2018) incorporates the feature of observing the opponent's (first-mover's) intentions. By introducing a passive third player—to whose beliefs the second mover can condition his choice—we are able to test whether observing the intentions of another player is a pre-requisite to the influence of vulnerability on the second-mover's behavior, and ultimately on triggering his guilt aversion.

Third, the within-game comparison allows us to disentangle the triggering effect of the opponent's vulnerability on the second-mover's distributional preferences (altruism) vs. belief-dependent preferences (guilt aversion), and check whether and how his sensitivity to each of the two types of social preference vary across games (*i.e.*, across types of vulnerability) and/or across treatments (*i.e.*, across types" of beliefs).

To provide theory-driven experimental hypotheses, we put forward a portable model of lexicographic altruism and role-dependent guilt, where both the first mover and the second mover can be altruistic toward the most disadvantaged player, while only the second mover can feel guilty, with his guilt sensitivity being mainly triggered by his role in the game. Therefore, he can feel guilty even if the opponent is not at all vulnerable and/or he cannot observe the opponent's intentions (e.g., the opponent is just an external observer).

Our experimental results show no significant difference in the proportion of guilt-averse second movers neither across games nor across treatments. This lack of significant difference suggests that vulnerability does not modulate guilt aversion of the second mover in our quasi-trust minigames. We interpret the insensitivity of guilt aversion to the co-player's vulnerability as further support to guilt being role-dependent in asymmetric games. More precisely, in line with Attanasi et al. (2016), we find that guilt aversion is essentially triggered by the specific role in the game. In particular, guilt is triggered by the role of second mover in two-stage dynamic games where the first mover manifests her intentions through his first-stage actions. Our results show that the second mover's guilt aversion is independent from the specific intentions of the first mover. It is rather activated by the dynamic structure of the game itself, *i.e.*, by the first mover's opportunity to express any intentions before the second mover's action. That opportunity leads the second mover to feel guilty also toward a passive player formulating beliefs about the first-mover's intentions.

The remainder of the paper is organized as follows. Section 2 presents in detail our four (new) games. Section 3 introduces the theoretical model and behavioral predictions. Section 4 describes the experimental design. Section 5 presents the results. Section 6 concludes.

2 Experimental Games

We introduce four games: the Investment game, the Reversed-Investment game, the Donation game and the Exploitation game. All the games have the same structure (see Figure 1).

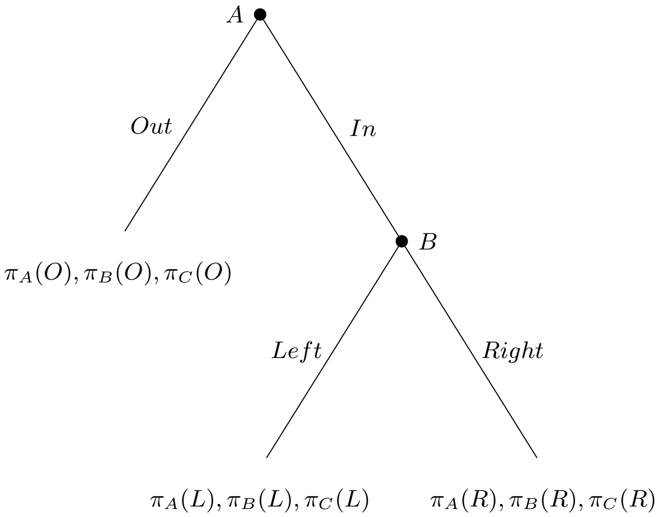


Figure 1: Common structure of the four games

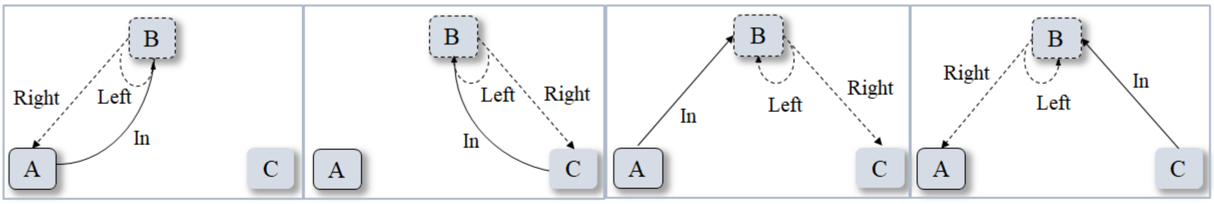
Our games are three-player games, with $i \in \{A, B, C\}$. Let players A and C be female ("she") and player B be male ("he"). Player A is the first mover, with her possible strategies $s_A \in \{In, Out\}$. Player B is the second mover – conditional on player A choosing *In* –, with his possible strategies $s_B \in \{Left \text{ if } In, Right \text{ if } In\}$. We denote player A's strategies *In* and *Out* with, respectively, I and O. We denote player B's strategies *Left* after *In* and *Right* after *In* with, respectively, L and R. Player C is always passive. Denote with O, L and R the terminal node after, respectively, (terminal) history O, (I, L) and (I, R). With this, we let $\pi_i(z)$ be the material payoff of player i at the terminal node $z \in \{O, L, R\}$.

If player A chooses *Out*, the game ends with material payoffs $\pi_A(O) = 170, \pi_B(O) = 100, \pi_C(O) = 30$.[3] These material payoffs can be interpreted as the distribution of initial endowments among the three players (A, B, and C). If player A chooses *In*: she sends 25 ECU to player B, with this amount being taken either from player A's endowment or player C's endowment, *depending on the game*; player B decides how to allocate these 25 ECU between himself and another player, the latter being either player A or player C, *depending on the game*. If player B chooses *Right*, he transfers the 25 ECU to the other player. If player B chooses *Left*, he transfers 5 ECU to the other player and keeps 20 ECU for himself. Each ECU transferred by player B to another player (A or C, depending on the game) is doubled.

Note that, given the terminal node $z \in \{O, L, R\}$, the payoff of player B, $\pi_B(z)$, is the same independently from the four games. This is because it does not depend on whether the ECU eventually sent by player A are taken from player A's or C's endowment. Neither it depends on whether the ECU transferred by player B go to player A or C. Therefore, if B chooses *Right* after *In*, his material payoff

[3] All material payoffs are expressed in Experimental Currency Unit (ECU) where 10 ECU = €1 (see the experimental procedures in Section 4.3).

is $\pi_B(R) = \pi_B(O)$; if he chooses *Left* after In, his material payoff is $\pi_B(L) = \pi_B(O) + 20$. As Figure 2 shows, the payoff manipulation across the four games concerns player A's and C's payoffs at terminal nodes $z \in \{L, R\}$.



	Investment	Reversed Investment	Donation	Exploitation
A	Vulnerable ex-ante & ex-post	Not Vulnerable	Vulnerable ex-ante	Vulnerable ex-post
C	Not vulnerable	Vulnerable ex-ante & ex-post	Vulnerable ex-post	Vulnerable ex-ante

Figure 2: Structure of the four games

Figure 3 presents the **Investment game**, where player A can entrust player B with 25 ECU from his endowment. Player B can then decide how to allocate these 25 ECU between player A and himself. In this game, player A is both ex-ante and ex-post vulnerable whereas player C is not vulnerable. Indeed, the Investment game is a simplified version (mini-game) of the classical Trust game (see Berg et al., 1995; Buskens and Raub, 2013; Attanasi et al., 2016), with the additional feature of an external observer, player C, who is not affected by the trustor and trustee's actions.

Figure 4 presents the **Reversed-Investment game** where player A can entrust player B with 25 ECU from player C's endowment. Player B can then decide how to allocate these 25 ECU between player C and himself. In this game, player A is not vulnerable and player C is both ex-ante and ex-post vulnerable. Thus, the Reversed-Investment game is a modified version of the Investment game where all monetary consequences of A's investment choice fall on C's payoff: A invests C's endowment and the doubled amount can enrich C.

Figure 5 presents the **Donation game** where player A can entrust player B with 25 ECU from his endowment. Player B can then decide how to allocate these 25 ECU between player C and himself. In this game, player A is ex-ante vulnerable and player C is ex-post vulnerable. Thus, the Donation game is a modified version of the Investment game where only the positive monetary consequences of A's investment choice fall on C's payoff: A invests her endowment and the doubled amount can enrich C. This is the Embezzlement game of Attanasi et al. (2019b).

Figure 6 presents the **Exploitation game** where player A can entrust player B with 25 ECU from player C's endowment. Player B can then decide how to allocate these 25 ECU between player A and himself. In this game, player A is ex-post vulnerable and player C is ex-ante vulnerable. Thus, the Donation game is a modified version of the Investment game where only the negative monetary consequences of A's investment choice fall on C's payoff: A invests C's endowment and the doubled amount can enrich A. This is the opposite of the Donation game.

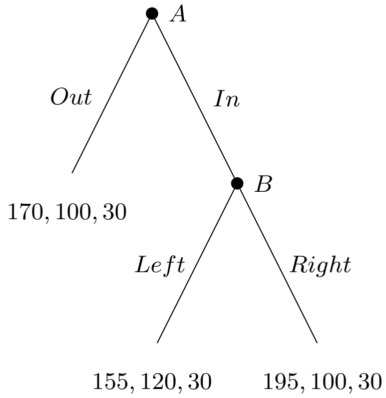


Figure 3: Investment game

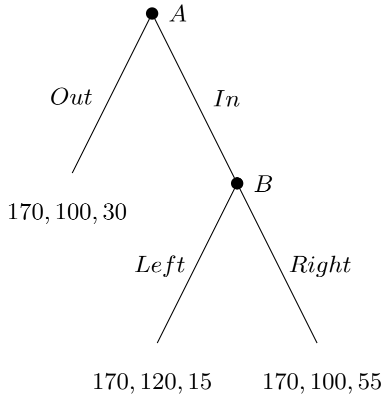


Figure 4: Reversed-Investment game

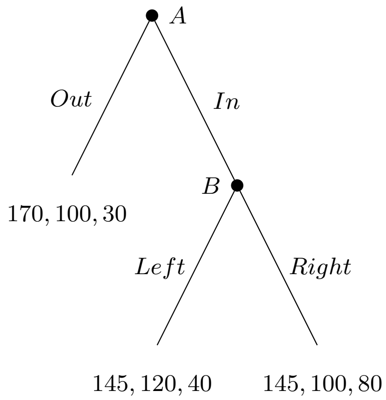


Figure 5: Donation game

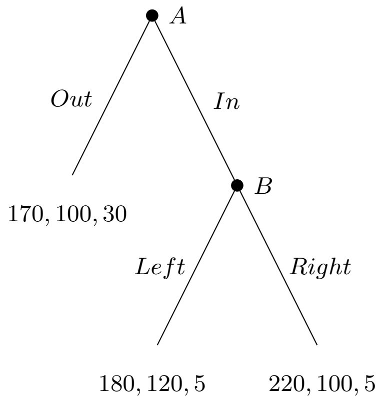


Figure 6: Exploitation game

Three game-independent features of the final distributions of material payoffs are worth noting. First, no decision can lead to the equalization of payoffs between two or three players. Hence, no payoff distribution should be more salient than the others. Second, the ranking of payoffs cannot be affected by the players' decisions. By doing so, we limit social comparison motives. Third, the total surplus at a given terminal node $z \in \{O, L, R\}$ is the same across games, keeping efficiency concerns constant across games.

3 The model

We elaborate our behavioral hypotheses relying on best-reply analysis rather than on Bayesian equilibrium. Indeed, a standard equilibrium analysis has no compelling foundation for games played one-shot (like ours) and in experiments on other-regarding preferences (see Section 6.2 of Attanasi et al., 2016).

3.1 Utility Functions

Figures 3-6 show respectively A's, B's and C's material payoffs (π_j, with $j \in \{A, B, C\}$) at each terminal node $z \in \{O, L, R\}$ of the games, *i.e.*, for each terminal history *Out* and (*In*, *Left*) and (*In*, *Right*).

First of all, since C is passive, we assume that she is purely self-interested. Therefore, **C's utility function** coincides with her material payoff, *i.e.*, $U_C(z) = \pi_C(z)$ for each $z \in \{O, L, R\}$. This is also motivated by the fact that, in each game and for each terminal history, C always gets the lowest payoff *i.e.*, $\min_j \pi_j(z) = \pi_C(z)$ at each terminal node $z \in \{O, L, R\}$..

As for **A's utility function**, we assume that she can be altruistic toward both B and C, since A always gets the highest payoff independently from the game and the strategy profile in that game, *i.e.*, $\max_j \pi_j(z) = \pi_A(z)$ at each terminal node $z \in \{O, L, R\}$. We also assume that A's altruistic preferences are lexicographic. More precisely, since C is always the most disadvantaged player and B is always the second most disadvantaged player, A is altruistic only toward C when C's payoff depends on A's strategy, and only toward B when C's payoff does not depend on A's strategy and B's payoff does depend on A's strategy.[4] Note that the latter occurs only in the Investment game, where C's payoff is the same at each terminal node.

We model A's feeling of altruism toward player $h \in \{B, C\}$, F_{Ah}, as A's utility derived from an increase in the payoff of player h. It is the product of two terms: $\phi_{Ah} \geq 0$, A's altruism sensitivity toward player h and $\pi_h(z)$, h's material payoff. A's utility is composed of her material payoff and her feeling of altruism toward player $h \in \{B, C\}$ (Eq. (1)):

$$U_A(\phi_{Ah}, z) = \pi_A(z) + F_{Ah}(\phi_{Ah}, z) \tag{1}$$

$$\text{where } F_{Ah}(\phi_{Ah}, z) = \phi_{Ah} \cdot \pi_h(z) \tag{2}$$

with $h = B$ in the Investment game and $h = C$ in the remaining three games. In each game, A's altruism toward player h depends on her belief about B's action after *In*. Let us define A's expected altruism toward player h as the difference between her altruism when she chooses In and her altruism when she chooses Out:

$$\mathbb{E}_A[F_{Ah}(\phi_{Ah}, z)|In] - \mathbb{E}_A[F_{Ah}(\phi_{Ah}, z)|Out] = \phi_{Ah} \cdot [\alpha_{AB} \cdot \pi_h(R) + (1 - \alpha_{AB}) \cdot \pi_h(L)] - \pi_h(O) \tag{3}$$

where α_{AB} is A's first-order belief that B chooses *Right* after *In*. Note that, in the Investment game (Figure 3), Eq. (3) is decreasing in α_{AB}; in the Reversed-Investment and the Donation game (Figures

[4]The latter assumption is broadly consistent with inequity aversion models (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999), since C is the most disadvantaged player.

4-5), Eq. (3) is increasing in α_{AB}; in the Exploitation game, Eq. (3) is negative and independent of α_{AB}.

Let us finally introduce **B's utility function** (Eq. (4)). Besides B's concern for his payoff π_B, we assume that B has lexicographic altruistic preferences toward disadvantaged players modeled as player A's ones (Eq. (5). Since C is always the most disadvantaged player and A is always the most advantaged player, B is altruistic only toward player C when C's material payoff depends on B's strategy, and is not altruistic when C's material payoff does not depend on B's strategy. Therefore, B can be altruistic toward player C in the Reversed-Investment and Donation games, and he cannot be altruistic toward any player in the Investment and Exploitation games.

Furthermore, in line with the role-dependent guilt model of Attanasi *et al.* (2016), we assume that B can feel guilty due to his role in the game (while A does not).[5] B's feeling of guilt, G_{Bjk}, with $j, k \in \{A, C\}$, represents his disutility derived from letting down j's beliefs on the strategy he will select, which will affect k's payoff, with j not necessarily equal to k. More precisely, it is the product of two terms: $\gamma_{Bjk} \geq 0$, B's guilt sensitivity about j's beliefs when B's strategy affects k's payoff; and the difference, if positive, between j's beliefs about k's payoff after In, $\mathbb{E}_j[\pi_k(z|In)]$, and k's actual material payoff after In, $\pi_k(z|In)$. Defining with α_{jB} j's first-order belief that B chooses *Right* after *In*, if $\mathbb{E}_j[\pi_k(z|In)] = \alpha_{jB} \cdot \pi_k(R) + (1 - \alpha_{jB}) \cdot \pi_k(L) > \pi_k(z|In)$, then B feels guilty from letting down j's beliefs on k's payoff. With this, B's utility after In is represented in Eq. (4), for $j, k \in \{A, C\}$:

$$U_B(\phi_{BC}, \gamma_{Bjk}, \alpha_{jB}, z|In) = \pi_B(z|In) + F_{BC}(\phi_{BC}, z|In) - G_{Bjk}(\gamma_{Bjk}, \alpha_{jB}, z|In) \tag{4}$$

$$\text{where } F_{BC}(\phi_{BC}, z|In) = \phi_{BC} \cdot \pi_C(z|In) \tag{5}$$

$$\text{and } G_{Bjk}(\gamma_{Bjk}, \alpha_{jB}, z|In) = \gamma_{Bjk} \cdot \max\{0, \mathbb{E}_j[\pi_k(z|In)] - \pi_k(z|In)\} \tag{6}$$

As for B's altruism in Eq. (5), it is $\phi_{BC} = 0$ in both the Investment and the Exploitation game. In the remaining two games, it is $F_{BC}(\phi_{BC}, R) > F_{BC}(\phi_{BC}, L)$, *i.e.*, *Right* after *In* is a more altruistic strategy than *Left* after *In*.

As for B's guilt aversion in Eq. (6), two clarifications are in order about $\gamma_{Bjk} \geq 0$ in each of the four games in Figures 3-6. First, for each of them, we implement an experimental design where the impact of each guilt sensitivity (toward player A and toward player C) is analyzed separately (see Section 4). In fact, we use a between-subject design to elicit B's belief-dependent strategy conditional on either

[5] 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, 2019a, 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 of emotion. Similar arguments can be provided in support of sensitivity to guilt being triggered only when playing in the role of player B (and not in the role of player A) in our four "trust" games of Figures 3-6.

A's (A treatment) or C's (C treatment) first-order beliefs about *Right* if *In*. Therefore, we make the auxiliary assumption that B's guilt is activated not by the material payoff of the player who is affected by B's strategy (*i.e.*, ex-post vulnerability), but by the belief of the player to whom B's strategy is exogenously conditioned. Therefore, we assume that $\gamma_{BC} = 0$ in the A treatment and $\gamma_{BA} = 0$ in the C treatment independently from the game, *i.e.*, independently from C's vulnerability and A's vulnerability, respectively.

Second, and in line with the previous clarification, for G_{Bjk} in Eq. (6), we extend Battigalli and Dufwenberg's (2007) definition of simple guilt as player B's disutility from letting down player j's expectations about *her own* material payoff, in the same direction as Attanasi et al. (2019b). In a series of Donation games comparable to ours, Attanasi et al. (2019b) have theoretically defined and experimentally measured G_{BAC} as B's disutility from letting down A's beliefs about C's material payoff, thereby being the first to show that "guilt towards another player can be triggered even when decisions have no direct consequences for that player" (Dufwenberg and Patel, 2019, p. 3), *i.e.*, absent ex-post vulnerability. Note that in the Donation game, the disappointed player A is ex-ante vulnerable.

Here we go several steps further in detecting a player's guilt aversion triggered by disappointing a second player's beliefs about a third player's payoff. In fact, we measure such guilt aversion in different three-player strategic settings where the disappointed player may or may not signal her intentions (resp., the active player A or the passive player C), or use her or the third player's payoff to entrust the guilt-averse player (*i.e.*, in the absence of both ex-post and ex-ante vulnerability of the disappointed player). This provides a multidimensional robustness test to the theoretical extension in Attanasi et al. (2019b) in the direction suggested by Attanasi et al. (2013) of B's guilt aversion mainly triggered by his own role.

More precisely, in the A treatment, where B's strategy is elicited conditional on A's first-order beliefs (*i.e.*, $j = A$ in Eq. (6)), we have the standard definition of Battigalli and Dufwenberg (2007) – with $k = A$ and so G_{BAA} – in two different strategic settings (Investment game and Exploitation game), and the extended definition – with $k = C$ and so G_{BAC} – in other two different strategic settings (Reversed-Investment game and Donation game). Correspondingly, in the C treatment, where B's strategy is elicited conditional on C's first-order beliefs (*i.e.*, $j = C$ in Eq. (6)), we have the standard definition of Battigalli and Dufwenberg (2007) – with $k = C$ and so G_{BCC} – in two different strategic settings (Reversed-Investment game and Donation game), and the extended definition – with $k = A$ and so G_{BCA} – in other two different strategic settings (Investment game and Exploitation game). Each pair of two games with the same vector (j, k), which disentangles whether the disappointed player j is or is not ex-post vulnerable, allows within-pair comparison where j is or is not ex-ante vulnerable. This leads to the four combinations of no vulnerability, only ex-ante vulnerability, only ex-post vulnerability and both ex-ante and ex-post vulnerability of the disappointed player, that we re-organize in Table 1 by looking at the between-game payoff manipulations of Figure 2.

		Ex-post Vulnerability	
		No	Yes
Ex-Ante Vulnerability	No	A in Rev. Investment C in Investment	A in Exploitation C in Donation
	Yes	A in Donation C in Exploitation	A in Investment C in Rev. Investment

Table 1: Vulnerability of disappointed players A and C in the four games

Note that, as in Battigalli and Dufwenberg (2007), B can feel guilty toward players being ex-post vulnerable (A in the Investment and Exploitation game; C in the Reversed-Investment and Donation game). As in Attanasi et al. (2019b), our model also assumes that B can feel guilty toward players being ex-ante vulnerable (A in the Investment and Donation game; C in the Reversed-Investment and Exploitation game). Furthermore, it also assumes that B can feel guilty toward players not being vulnerable at all (A in Reversed-Investment and C in the Investment game), if B is aware that those players have formed expectations about his own behavior. Due to his role in the game, after observing A's In choice, B dislikes disappointing A's beliefs α_{AB} in the A treatment ($\gamma_{BAk} \geq 0$) and C's beliefs α_{CB} in the C treatment ($\gamma_{BCk} \geq 0$) regardless of the ex-ante - ex-post vulnerability combinations of Table 1. Finally, note that our model is silent about the size of B's guilt sensitivity γ_{Bjk} in each treatment-game combination: it only assumes that it can be $\gamma_{Bjk} > 0$ for each (j, k).

3.2 Best-reply analysis

We now analyze the best-reply strategies of A and B as a function, for the former, of her altruism toward C, and for the latter, of his altruism toward and his guilt toward A (in treatment A) and toward C (in treatment C). We do it separately for each of the four games, and in a separate subsection for each active player in the games.

3.2.1 Player A's best-reply functions

In each of the four games, we define A's best-strategy as a function of her first-order belief α_{AB} that B chooses $Right$ after In and of her sensitivity to altruism toward player B, ϕ_{AB}, in the Investment game, and toward player C, ϕ_{AC}, in the other three games.

Investment Game. In the Investment game, C's payoff is unaffected by A's strategy (C only is an external observer) while B's payoff is affected. Therefore, by lexicographic altruism, A is only altruistic toward B. A's expected altruism toward player B from Eq. (3) is $20 \cdot \phi_{AB} \cdot (1 - \alpha_{AB})$, hence decreasing in A's belief that B would choose *Right* after In. A's expected material payoff if she chooses In, $\mathbb{E}_A[\pi_A(z|In)] = \alpha_{AB} \cdot 195 + (1 - \alpha_{AB}) \cdot 155$, is increasing in A's belief that B would choose *Right* after In, and greater than the material payoff if she chooses *Out*, $\pi_A(O) = 170$, if $\alpha_{AB} > 3/8$. Putting together A's material and altruistic interest, we find that if $\alpha_{AB} > 3/8$ then A chooses In whatever her altruism sensitivity ϕ_{AB}. For lower first-order beliefs about *Right* after In, A chooses In only if $\phi_{AB} > 0$, and, given $\alpha_{AB} < 3/8$, she chooses In more the higher her altruism sensitivity ϕ_{AB}. To summarize: a selfish or moderately altruistic A chooses In more the higher her first-order belief that B chooses *Right* after *In*; a highly altruistic A chooses In regardless of her first-order belief. Therefore, given α_{AB}, the higher ϕ_{AB}, the higher the likelihood that A chooses In.

Reversed-Investment Game. In this game A's material payoff is the same at each terminal node. Therefore, to state A's best-reply function, we only rely on her expected altruism toward player $h = C$ from Eq. (3): A chooses In if $\phi_{AC} \cdot (40 \cdot \alpha_{AB} - 15) > 0$. Thus, an altruistic player A (i.e, with $\phi_{AC} > 0$) with high enough first-order belief of *Right* after In (*i.e.*, $\alpha_{AB} > 3/8$) chooses In. This holds independently from her sensitivity to altruism. To summarize: a selfish A is indifferent between *In* and *Out* for each first-order belief α_{AB}; any altruistic player A chooses In only for high enough first-order belief of *Right* after In.

Donation Game. In this game the In choice decreases A's material payoff and increases C's expected material payoff. Therefore, to find A's best-reply function, we have to compare the certain decrease in A's material payoff when she chooses In, *i.e.*, $\pi_A(z|In) - \pi_A(O) = 25$, with her expected altruism toward player $h = C$ from Eq. (3), which is increasing in her first-order belief of *Right* after In. With this, we find that A chooses *In* if $\alpha_{AB} > 3/8 \cdot (1/\phi - 2/3)$. Thus, a necessary condition for choosing In is altruism toward player C, *i.e.*, $\phi_{AC} > 0$. But this is not sufficient: A's first-order belief of *Right* after In must be high enough, less high for higher ϕ_{AC}. To summarize: a selfish or moderately altruistic A never chooses *In* regardless of her first-order belief α_{AB}; a highly altruistic A chooses In only for high enough α_{AB}.

Exploitation Game. In this game A's altruism if she chooses In does not depend on her first-order belief of *Right* after In, since C gets the same material payoff at each of the two terminal nodes $z|In$. Therefore, expected altruism toward player $h = C$ from Eq. (3) shrinks to $-25 \cdot \phi_{AC}$, *i.e.*, differently from the other three games, the altruistic action is Out. A's increase in the material payoff if she chooses In, $\mathbb{E}_A[\pi_A(z|In)] - \pi_A(O)]$, positively depends on α_{AB}. Putting together A's material and altruistic interest, we find that A chooses In if $\alpha_{AB} > 3/8 \cdot (\phi - 2/3)$. Thus, a sufficient condition for choosing In

is selfishness, *i.e.*, $\phi_{AC} = 0$. But this is not necessary: also a moderately altruistic A (*i.e.*, $\phi_{AC} < 2/3$) or a highly-altruistic A with high first-order belief of *Right* after *In* choose *In*. To summarize: a selfish or moderately altruistic A always chooses *In* regardless of her first-order belief α_{AB}; a highly-altruistic A chooses *In* only for high enough α_{AB}.

In Figure 7 we summarize the best-reply strategies of A players for different altruistic types ϕ_{Ah} and for the four possible first-order beliefs (about player B choosing *Right* after *In*) that A players can hold in our experiment, *i.e.*, $\alpha_{AB} \in \{0, 1/3, 2/3, 1\}$, as we will explain in Section 4.

α_{AB}	0				1/3				2/3				1			
φ_{Ah}	Inv	Rev Inv	Don	Exp	Inv	Rev Inv	Don	Exp	Inv	Rev Inv	Don	Exp	Inv	Rev Inv	Don	Exp
0 (selfish)	Out	In	Out	In	Out	In	Out	In	In	In	Out	In	In	In	Out	In
(0.00, 0.13]	Out	Out	Out	In	Out	Out	Out	In	In	In	Out	In	In	In	Out	In
(0.13, 0.30]	Out	Out	Out	In	In	Out	Out	In	In	In	Out	In	In	In	Out	In
(0.30, 0.41]	Out	Out	Out	In	In	Out	Out	In	In	In	Out	In	In	In	In	In
(0.41, 0.64]	Out	Out	Out	In	In	Out	Out	In	In	In	In	In	In	In	In	In
(0.64, 0.67]	Out	Out	Out	In	In	Out	In	In	In	In	In	In	In	In	In	In
(0.67, 0.75]	Out	Out	Out	Out	In	Out	In	In	In	In	In	In	In	In	In	In
(0.75, 1.55]	In	Out	Out	Out	In	Out	In	In	In	In	In	In	In	In	In	In
(1.55, 2.44]	In	Out	Out	Out	In	Out	In	Out	In	In	In	In	In	In	In	In
(2.44, 3.33]	In	Out	Out	Out	In	Out	In	Out	In	In	In	Out	In	In	In	In
(3.33, +∞)	In	Out	Out	Out	In	Out	In	Out	In	In	In	Out	In	In	In	Out

Figure 7: A's predicted behavior in the four games, depending on her altruistic sensitivity (ϕ_{Ah}) and first-order belief about *Right* after *In* (α_{AB}). (A's altruistic strategy in dark blue and A's selfish strategy in light blue color)

Note that due to the assumption of lexicographic altruistic preferences, for ϕ_{Ah}, $h = B$ in the Investment game and $h = C$ in the other three games. Also note that we used the dark blue color for A's altruistic strategy and the light blue for A's selfish one. Thus, A's welfare-maximizing strategy *In* is in dark-blue color in the first three games and in light-blue in the Exploitation game, where the altruistic strategy is *Out*. Finally, for a selfish player A in the Reversed-Investment game (where she is indifferent between *In* and *Out* for each α_{AB}) we assume that she chooses *In*. The way we break the tied strategy is motivated by experimental demand effects due to both welfare maximization (see, e.g., Charness and Rabin, 2002) and to the fact that the *In* action lets the game unfold with B's strategy being payoff-relevant for himself and player C.

3.2.2 Player B's best-reply functions

In each of the four games in Figures 3-6, if player B chooses *Right* after *In*, he entirely transfers to another player the amount of money that the *In* choice of player A has entitled him to manage. If instead he chooses *Left* after *In*, he only transfers a small portion (20%) of that amount. Therefore, relying on

Eqs. (4–6), we define for each treatment (A and C) player B's *Willingness-to-Transfer function* (WT) as the difference between his utility from *Right* and his utility from *Left* after In. Both these terms are expected utilities since player B forms beliefs about the first-order beliefs α_{jB} of the co-player j toward whom he feels guilty ($j = A$ in the A treatment and $j = C$ in the C treatment).[6] These are his conditional second-order beliefs $\beta_{Bj} = \mathbb{E}_B[\alpha_{jB}|In]$, *i.e.*, for $j \in \{A, C\}$, conditional on player A choosing In.[7] After In, the more player B prefers to choose *Right* rather than *Left*, the higher his willingness to transfer the money that the In choice of player A has entitled him to manage:

$$\begin{aligned} WT(\phi_{BC}, \gamma_{Bjk}, \alpha_{jB}, z|In) &= \mathbb{E}_B[U_B(\phi_{BC}, \gamma_{Bjk}, \alpha_{jB}, R)] - \mathbb{E}_B[U_B(\phi_{BC}, \gamma_{Bjk}, \alpha_{jB}, L)] \\ &= \pi_B(R) - \pi_B(L) + \phi_{BC}[\pi_C(R) - \pi_C(L)] + \gamma_{Bjk}\beta_{Bj}[\pi_k(R) - \pi_k(L)] \quad (7) \end{aligned}$$

Note that, given the common structure, in each of the four games of Figures 3-6, it is $\pi_B(R)-\pi_B(L) = -20$. With this, for each of them we can find player B's best-reply strategy as a function of his sensitivity to altruism toward player C, ϕ_{BC}, her second-order belief β_{jB} that he will choose *Right* after In, and her sensitivity γ_{Bjk} to guilt toward the (first-order belief α_{Bj} of the) player on whom B's second-order belief relies. In particular, a rational (best-replying) player B of type $(\phi_{BC}, \gamma_{Bjk})$ and belief β_{jB} chooses *Right* after In if $WT > 0$ in Eq. (7) and *Left* otherwise.

Investment Game. By construction, $\pi_C(R)-\pi_C(L)$, thus $F_{BC} = 0$ in Eq. (5), *i.e.*, player B cannot be altruistic toward player C. Furthermore, A is ex-post vulnerable, hence $k = A$ in Eq. (7). With this, WT in Eq. (7) reduces to:

$$\pi_B(R) - \pi_B(L) + \gamma_{BjA} \cdot \beta_{Bj} \cdot [\pi_A(R) - \pi_A(L)] \quad (8)$$

By substituting the game payoffs of Figure 3, Eq. (8) becomes $-20 + 40 \cdot \gamma_{BjA} \cdot \beta_{Bj}$, which is strictly positive for all type-belief pairs $(\gamma_{BjA}, \beta_{Bj})$ such that $\gamma_{BjA} \cdot \beta_{Bj} > 1/2$. Therefore, a guilt-averse B chooses *Right* after In more for greater guilt sensitivity γ_{BjA} and greater conditional second-order belief β_{Bj} of *Right* after In. According to our model of Eqs. (4-6), the previous relation holds in both in the A treatment, *i.e.*, for type-belief pairs $(\gamma_{BAA}, \beta_{BA})$, and in the C treatment, *i.e.*, for $(\gamma_{BCA}, \beta_{BC})$.

Reversed-Investment Game. In this game, C is ex-post vulnerable, hence $k = C$ in Eq. (7),

[6] In each of the four games, we assume that player B best-responds *as if* he had truly observed player A's move. This holds by standard expected-utility maximization, except for the cases where player B is certain that player A has chosen *Out*. Thus, we need the additional assumption that player B has a belief conditional on *In*, even if he is certain of *Out*. Indeed, in our experiment (see Section 4) player B's decision is made under the strategy method (*i.e.*, both when player A has chosen *Out* and when she has chosen *In*).

[7] More precisely, we reason as if player B has a point belief β_{Bj} about α_{jB} conditional on *In*.

which becomes:

$$\pi_B(R) - \pi_B(L) + (\phi_B + \gamma_{BjC} \cdot \beta_{Bj}) \cdot [\pi_C(R) - \pi_C(L)] \tag{9}$$

By substituting the game payoffs of Figure 4, Eq. (9) becomes $-20 + 40 \cdot (\phi_{BC} + \gamma_{BjC} \cdot \beta_{Bj})$, which is strictly positive for all type-belief pairs $((\phi_{BC}, \gamma_{BjC}), \beta_{Bj})$ such that $\phi_{BC} + \gamma_{BjC} \cdot \beta_{Bj} > 1/2$. Therefore, a guilt-averse B chooses *Right* after In more for greater guilt sensitivity γ_{BjC} and greater conditional second-order belief β_{Bj} of *Right* after In. According to our model of Eqs. (4-6), the previous relation holds both in the A treatment, *i.e.*, for type-belief pairs $(\gamma_{BAC}, \beta_{BA})$, and in the C treatment, *i.e.*, for $(\gamma_{BCC}, \beta_{BC})$. Furthermore, independently from the treatment, the higher ϕ_{BC}, B's sensitivity to altruism toward player C, the lower the guilt sensitivity γ_{BjC} and the second-order belief β_{Bj} required for player B to choose *Right* after In. Finally, for high enough sensitivity to altruism (*i.e.*, $\phi_{BC} > 1/2$), player B chooses *Right* after In regardless of his second-order belief β_{Bj}.

Donation Game. In this game, C is ex-post vulnerable, hence $k = C$ in Eq. (7). Therefore, WT in this game is the same as in Eq. (9). By substituting the game payoffs of Figure 5, given the similar structure between the Reversed-Investment and the Donation game (*i.e.*, $\pi_B(R) - \pi_B(L) = -20$ and $\pi_C(R) - \pi_C(L) = 40$), we find the same subset of type-belief pairs $((\phi_{BC}, \gamma_{BjC}), \beta_{Bj})$ for which Eq. (9) is strictly positive, *i.e.*, $\phi_{BC} + \gamma_{BjC} \cdot \beta_{Bj} > 1/2$. Therefore, and again independently from the treatment $j = A$ or $j = C$, the same considerations made for the Reversed-Investment game hold about the positive impact of each parameter of the vector $((\phi_{BC}, \gamma_{BjC}), \beta_{Bj})$ over the likelihood of player B choosing *Right* after In.

Exploitation Game. By construction, $\pi_C(R) - \pi_C(L)$, thus $F_{BC} = 0$ in Eq. (5), *i.e.*, player B cannot be altruistic toward player C. Furthermore, A is ex-post vulnerable, hence $k = A$ in Eq. (7). Therefore, WT in this game is the same as Eq. (8). By substituting the game payoffs of Figure 6, given the similar structure between the Investment and the Exploitation game (*i.e.*, $\pi_B(R) - \pi_B(L) = -20$ and $\pi_A(R) - \pi_A(L) = 40$), we find the same subset of type-belief pairs $(\gamma_{BjA}, \beta_{Bj})$ for which Eq. (8) is strictly positive, *i.e.*, $\gamma_{BjA} \cdot \beta_{Bj} > 1/2$. Therefore, and again independently from the treatment $j = A$ or $j = C$, the same considerations made for the Investment game hold about the positive impact of each parameter of the vector $(\gamma_{BjA}, \beta_{Bj})$ over the likelihood of player B choosing *Right* after In.

3.3 Behavioral hypotheses

First, we state the behavioral hypotheses for A-subjects. The first two hypotheses refer to behavior in each of the four games analyzed separately, and are mainly meant to test whether our assumption of

A's lexicographic altruism is verified in our sample. All the others refer to comparative behavior across games, and are meant to show that the motivation behind the In choice of player A—which triggers B's guilt in the subgame where the latter is active—is different, and sometimes opposite, in different games. This motivation depends on A's and/or C's type of vulnerability, if any, which vary across games.

Given A's sensitivity to altruism, the theoretical predictions in Figure 7 show a positive relation between the likelihood of the In choice and A's first-order belief of *Right* after *In*. This holds regardless of the game. In particular, in the Investment game it holds regardless of A's sensitivity to altruism, in the Reversed-Investment game for all altruistic subjects, in the Donation game only for highly-altruistic subjects, and in the Exploitation game for all subjects but highly-altruistic ones. Considering heterogeneity in A's sensitivity to altruism, we elaborate a hypothesis about A's belief-dependent behavior in each of the four games.

H.A.1 [Choice-belief correlation] The frequency of *In* choices by A-subjects increases in their first-order belief about B-subjects choosing *Right* in each of the four games (Investment, Reversed-Investment, Donation and Exploitation).

Given A's first-order belief of *Right* after *In*, the theoretical predictions in Figure 7 for the Investment and the Donation game also show a positive relation between the likelihood of the In choice and A's sensitivity to altruism. For the former, this only holds under first-order beliefs of a distrustful B ($\alpha \leq 1/3$). For the latter, the positive relation holds for every positive first-order belief. Conversely, for both the Reversed-Investment and the Exploitation game they show a negative relation between the likelihood of the In choice and A's sensitivity to altruism. For the former, this only holds under first-order beliefs of a distrustful B ($\alpha \leq 1/3$). For the latter, the negative relation holds for every first-order belief. Considering heterogeneity in A's sensitivity to altruism and in A's first-order belief of *Right* after *In*, we elaborate a hypothesis about A's altruistic type-dependent behavior in each of the four games.

H.A.2 [Choice-type correlation] The frequency of *In* choices by A-subjects increases in their sensitivity to altruism in both the Investment and the Donation game. It decreases in A-subjects' sensitivity to altruism in both the Reversed-Investment and the Exploitation game.

In the next two hypotheses, besides heterogeneity in A's sensitivity to altruism, we also make the operational assumption that this sensitivity does not vary too much across games within-subject (*i.e.*, each A-subject has a ϕ_{Ah} in the same interval of Figure 7 for each of the four games). This is required in order to elaborate between-game comparisons in terms of type-dependent behavior, given a low ($\alpha_{AB} \leq 1/3$) or a high ($\alpha_{AB} \geq 2/3$) first-order belief about B choosing *Right*, *i.e.*, given beliefs of a distrustful B or a trustful B, respectively.

As for A's choice under beliefs of a distrustful B, we elaborate the next hypothesis by looking at the first two panels of Figure 7. We take the Donation game as reference game (control), since it is the only one where it is possible to identify a subset of altruistic A-types predicted to choose *Out* in that game and *In* in another game. In fact, for $\alpha_{AB} = 0$, A-subjects are predicted to choose *Out* in the Donation game regardless of ϕ_{Ah}, while in the other three games there exists a subset of altruistic A-types predicted to choose *In*. These types are the highly-altruistic A-subjects in the Investment game, the selfish ones in the Reversed-Investment game, and the selfish and lowly-altruistic ones, considered together, in the Exploitation game. In the same line, for $\alpha_{AB} = 1/3$, we can identify a subset of A-types predicted to choose *Out* in the Donation game and *In* in another game. In the Investment game, this relation holds for lowly-altruistic A-subjects (*i.e.*, $0.13 \leq \phi_{Ah} \leq 0.64$). In the Reversed-Investment game, the same relation holds for the selfish A-subjects ($\phi_{Ah} = 0$). In the Exploitation game, it holds for selfish and lowly-altruistic A-subjects (*i.e.*, $\phi_{Ah} \geq 0.64$).

We combine the two sets of predictions for $\alpha_{AB} = 0$ and $\alpha_{AB} = 1/3$ in a unique hypothesis about game-dependent behavior of A-subjects who believe that B would be distrustful after *In*, *i.e.*, that he would more likely choose *Left* ($\alpha_{AB} \leq 1/3$)).

H.A.3 [Choice under beliefs of a distrustful B] For A-subjects thinking that *Left* is the most likely action of B-subjects, the frequency of In choices in the Donation game is lower than: (i) in the Reversed-Investment game for selfish types; (ii) in the Exploitation game for selfish and lowly-altruistic types; (iii) in the Investment game for lowly-altruistic and highly-altruistic types.

As for A's choice under beliefs of a trustful B ($\alpha_{AB} \geq 2/3$), we elaborate the next hypothesis by looking at the last two panels of Figure 7. For both $\alpha_{AB} = 2/3$ and $\alpha_{AB} = 1$, the predicted choice for both the Investment and the Reversed-Investment game is *In* regardless of the altruism sensitivity ϕ_{Ah}, while it depends on the altruism sensitivity in the other two games. In particular, in the Donation game *In* is predicted to be chosen by lowly-altruistic and highly-altruistic types ($\phi_{Ah} > 0.41$ for $\alpha_{AB} = 2/3$ and $\phi_{Ah} > 0.30$ $\alpha_{AB} = 1$), while in the Exploitation game *In* is chosen by all types with "reasonable" (*i.e.*, not extremely high) altruistic sensitivity (*i.e.*, $\phi_{Ah} < 2.44$ for $\alpha_{AB} = 2/3$ and $\phi_{Ah} < 3.33$ for $\alpha_{AB} = 1$). Absence of a significant difference in the frequency of *In* choices between the Exploitation game and each of the two Investment games would indirectly disclose absence of such types in our pool of A-subjects. With this, we present our last hypothesis about A's game-dependent behavior. We formulate it by taking the Investment game as reference game (control) since it is the closest to the standard version of the trust game in the literature, and it is directly comparable with the Donation game in H.A.3.

H.A.4 [Choice under beliefs of a trustful B] For A-subjects thinking that *Right* is the most likely action of B-subjects, the frequency of In choices in the Investment game is: (i) the same as in the Reversed-Investment game, regardless of the altruistic type; (ii) higher than in the Donation game for selfish and lowly-altruistic types; (iii) higher than in the Exploitation game for highly-altruistic types.

Note that we derived H.A.1-H.A.4 without specifying the treatment (A or C) since, in our experiment, players A are unaware, when they make their choices, of the treatment, *i.e.*, of the player (A or C) of which B considers the first-order beliefs for his belief-dependent strategy. Therefore, A's behavior should be treatment-independent.

Furthermore, and more importantly, the hypotheses about the behavior of A-subjects have been elaborated with a twofold aim. First, to highlight *belief-dependent behavior*. H.A.1-H.A.2 are meant to verify that in each of the four games A's choice to entrust B through *In* depends on the interplay between her altruistic-type ϕ_{Ah} and her belief-type α_{AB}: for some given level of A's sensitivity to altruism, A's choice is belief-independent; for any other level of it, A trusts B more through *In* the more she expects B to be trustworthy through *Right* after *In*. The latter is the channel for the activation of A's disappointment in the case of *Left* after *In*, which is the primitive of B's guilt aversion through his second-order belief of *Right* after *In*.

Second, we elaborate our hypotheses to underline *game-dependent behavior*. In fact, H.A.3-H.A.4 are meant to verify that the motivation behind A's *In*tention of entrusting B is game-dependent. In fact, in some games, a B player observing *In* should expect that A's intention is to increase her own payoff (e.g., in the Investment and Exploitation game), while in some others she wishes to increase C's one (e.g., in the Reversed-Investment and Donation game). Nonetheless, given A's first-order belief of *Right*, the same *In* choice in different games would signal a different sensitivity to altruism by the A player who made it. Therefore, intuitively, also B's guilt sensitivity should be game-dependent, if B would care about the different intentions behind A's *In* choice, which in turn stems from the different combinations of ex-ante and ex-post vulnerability that each game structure associate to roles A and/or C.

However, our study relies on the opposite intuition that B's *belief-dependent behavior* is *game-independent*. We postulate that in our four quasi-trust mini-games, B's guilt aversion is not triggered by A's specific intention in a given game. Rather, it is activated by the fact itself that A can express her intention to trust B (through the *In* choice) and that B can observe this choice. More precisely, B's guilt aversion is triggered by his role of second mover in a two-stage game with perfect information: this role is game-independent, and so it should be his sensitivity to guilt. This intuition allows us to go one step further: we also postulate that B's *belief-dependent behavior* is *treatment-independent*, *i.e.*, B's guilt aversion is triggered even though A's intention is mediated by another player's first-order beliefs, namely the passive player C. Therefore, given A's *In*tention of entrusting B, the elicited B's belief-dependent behavior should be the same regardless of whether it is related to A's or C's first-order beliefs (respectively, in treatment A and in treatment C).

Again, we have two families of hypotheses: those that refer to similarities in B's belief-dependent and type-dependent behavior in each of the four games separately (H.B.1-H.B.2), and those that refer to comparative behavior across games (H.B.3-H.B.5).

The first family of hypotheses has three specific aims. First, we check for positive correlation between the frequency of *Right* choices and second-order beliefs (H.B.1), and between the former and guilt sensitivities (H.B.2), in order to verify the activation of guilt aversion in a non-negligible fraction of B players.

Second, recall our assumption that in the two games where B's choice concerns the material payoff of player C ($k = C$ in Eq. (4)), B's altruism from Eq. (5) matters (i.e., in the Reversed-Investment and the Donation game), while it does not play any role in the Investment and Exploitation game ($k = C$ in Eq. (4). Therefore, H.B.2 is also meant to test that, while the choice-type correlation is game-independent for belief-dependent preferences (guilt sensitivity), it is game-dependent for belief-independent preferences (altruism sensitivity), the latter correlation being absent in the Investment and Exploitation game, where the poorer and passive player C is not ex-post vulnerable.

Third, we check for similar behavior in both treatments, *i.e.*, independently of the player (A or C) whose first-order beliefs are used to elicit B's second-order beliefs and belief-dependent strategies. Let us recall that, according to Battigalli and Dufwenberg (2007) and follow-up papers, B could feel guilty only in treatment-game combinations where his/her strategy and second-order beliefs are conditioned to the first-order beliefs of a player who is ex-post vulnerable. In our design, this only occurs in 4 out of the 8 treatment-game combinations, namely in the Investment game and Exploitation game under treatment A, and in the Reversed-Investment and Donation game under treatment C. According to our model of Eqs. (4) and (6), B's guilt aversion is also activated when his/her strategy and second-order beliefs are conditioned to the first-order beliefs of a player who is not ex-post vulnerable or even not vulnerable at all.[8] Note that all treatment comparisons are made at a between-subject level.

H.B.1 [Choice-belief correlation] The frequency of *Right* choices by guilt-averse B players increases in their second-order beliefs about *Right* in each of the four games. All this holds independently from the treatment.

H.B.2 [Choice-type correlation] Given a positive second-order belief, the frequency of *Right* choices of B players increases with: (i) their altruism sensitivity only in the Reversed-Investment and Donation game; (ii) their guilt sensitivity in each of the four games. All this holds independently from the treatment.

Finally, we present the second family of behavioral hypotheses for B-subjects, which consider the

[8] For the Exploitation game, B's guilt aversion in treatment C, where B's decision relies on the beliefs of the passive player from whom player A has taken the money to increase her own payoff, can also be interpreted as B's unwillingness to get C's money. In fact, the higher α_{CB}, the more C beliefs that B will not take any of her money by choosing *Right* after *In*.

four games and the two treatments together. The first hypothesis challenges the categorization of the four games in terms of vulnerability of disappointed players A and C, provided in Table 1. Our model is silent about this issue. It just assumes that B may feel guilt-averse even in absence of any type of vulnerability (*i.e.*, in the Reversed-Investment game toward A's beliefs, and in the Investment game toward C's beliefs) and independently from the vulnerability type (ex-ante vs. ex-post). Therefore, we elaborate a null hypothesis stating similar guilt-averse behavior across the four games at a within-subject level, given the treatment A (first rows of the four cells of Table 1) or C (second rows of the four cells of Table 1). In terms of disappointed player j's vulnerability, this means that, given $j \in A, C$, player B discloses the same guilt sensitivity regardless of whether j is not vulnerable, (only) ex-post vulnerable, (only) ex-ante vulnerable, or (both) ex-post & ex-ante vulnerable.

H.B.3 [Within-subject vulnerability-independent guilt] Within the same treatment, the fraction of elicited guilt-averse subjects is not significantly different across the four games. This holds in both treatment A and in treatment C.

The next hypothesis is at the core of our extension of Battigalli and Dufwenberg (2007), assessing that player B can feel guilty also when disappointing the beliefs of a player whose payoff is not affected by B's decision (*i.e.*, $j \neq k$ in Eq. (4). Our model is silent about the amount of such guilt. However, to elaborate H.B.4, we rely on the experimental results of Attanasi et al. (2019b), who tested this hypothesis in the Donation game, finding no significant difference in B's detected guilt between $j = k = C$ and $A = j \neq k = C$ in Eq. (4), *i.e.*, respectively, between C's ex-post and A's ex-ante vulnerability in Table 1. H.B.4 aims at extending the results of Attanasi et al. (2019b) in several dimensions. Besides replicating the result of B's guilt insensitivity to the type of vulnerability in the Donation game, it uses the other three pairwise comparisons of game-treatment combinations with $j = k$ vs. $j \neq k$ in Eq. (4) allowed by our design in order to test (see Table 1): ex-post vs. ex-ante vulnerability in the Exploitation game (treatment A vs. C); ex-ante & ex-post vs. no vulnerability in the Investment game (treatment A vs. C); ex-ante & ex-post vs. no vulnerability in the Reversed-Investment game (treatment C vs. A). Therefore, with respect to Attanasi et al. (2019b), besides comparing the effect of the two types of vulnerability with reversed roles of the disappointed player (ex-post for A and ex-ante for C), we also aim at stating whether, given a game structure, the trustee's guilt is not triggered at all by the fact that the disappointed player is vulnerable. This can be assessed by measuring the joint effect of the two types of vulnerabilities (*i.e.*, A in the Investment game or C in the Reversed-Investment game) taking as control a non-vulnerable disappointed player (*i.e.*, C in the Investment game or A in the Reversed-Investment game, respectively). [9]

[9]Note that H.B.4 also challenges some recent empirical findings in the literature of psychological games,*i.e.*, that signaling one's own intentions matter (see, e.g., Bellemare et al. (2017), detecting more guilt among trustees

H.B.4 [Between-subject vulnerability-independent guilt] Across the two treatments, the fraction of elicited guilt-averse subjects is not significantly different within the same game.

Taken together, H.B.1-H.B.2 postulate that guilt is activated in each of the eight game-treatment combinations, and H.B.3-H.B.4 posit the same fraction of guilt-averse B players across the four games and across the two treatments. Conversely, relying on the assumption of B's lexicographic altruism, H.B.2 asserts that B's altruism is only activated in the Reversed-Investment and Donation game, thereby boosting B's other-regarding behavior in these two games as compared to the Investment and Exploitation game. Being guilt aversion and altruism the only two other-regarding motivations of player B in Eq. (4), this would ultimately lead to a greater fraction of selfish B players (*i.e.*, always choosing *Left*) in the latter two games. Furthermore, being this game difference only due to a belief-independent preference (altruism), it should hold regardless of the treatment, since that the treatment manipulation only concerns the primitive of B's second-order beliefs. A successful test of H.B.5 would confirm our assumption of game-dependent altruism (*i.e.*, only triggered when C is ex-post vulnerable) as opposed to game-independent guilt (*i.e.*, triggered regardless of C's ex-post and/or ex-ante vulnerability).

H.B.5 [Game-dependent altruism] The fraction of elicited selfish subjects is significantly greater in the Investment and Exploitation game than in the Reversed-Investment and Donation game. This holds independently of the treatment.

Note the our predictions about player B's behavior, H.B.1-H.B.5, elaborated according to our model of guilt aversion in Eq. (4), are in contrast with the idea of vulnerability that we discussed in the Introduction of the paper when motivating similarities and differences among the four games in Figure 3-6. There we postulated that if the structure of the game makes a player vulnerable to another player's decision, then the decision maker's guilt aversion might increase, and this increase could depend on the dimension of vulnerability (ex-ante or ex-post). Our view, stated by H.B.1-H.B.5, is instead that guilt aversion is insensitive to disappointing payoff-vulnerable players or that at most vulnerability plays a secondary effect with respect to guilt driven by disappointing another player's beliefs. This assertion relies on the assumption that guilt can also be triggered when disappointing another player's beliefs about a third player's intentions, which is the key contribution of our theoretical extension of Battigalli and Dufwenberg (2007) provided by our model in Eq. (4)-(6).

than among dictators). Therefore, in each of the four aforementioned pairwise comparisons, B's guilt sensitivity should be greater toward a player signaling her intentions through her previous move (A, similar to a trustor in a trust game) than toward the passive player C (similar to a recipient in a dictator game).

4 Experimental Design

In our experimental design, each subject went through the four games of Figures 3-6: Donation, Investment, Reversed-Investment and Exploitation. The games were renamed with neutral labels ("North", "South", "East", and "West"). In each game, participants played in groups of three, with roles (A, B and C) assigned at the beginning of the experiment and maintained fixed in the four games. Groups were re-matched across games according to an absolute stranger protocol. We manipulated two dimensions. First, within-subject, we varied the order of presentation of the four games, which was randomized across experimental sessions. Second, between-subject, we varied the treatment: Treatment A and Treatment C. We now describe in details the experimental design and the procedures.

4.1 Decisions and beliefs

First-order belief elicitation We elicited, for each game, player B's and player C's first-order belief on the frequency of players A choosing *In*. They had to report, for each game, their belief about the number of players A, out of three players A randomly selected in the session, who choose *In* (four possible first-order beliefs: zero, one, two or three out of three players A). We also elicited, for each game, player A's and player C's first-order beliefs on the frequency of B-subjects choosing *Right* after *In*. Similarly, they had to report, for each game, their belief about the number of players B, out of three players B randomly selected in the session, who choose *Right* conditional on player A choosing *In* (four possible first-order beliefs). For each role, one belief out of the four elicited in the four games was randomly selected at the end of the session and paid €1 if accurate.

Player A's decision For each game, player A chose between *In* or *Out*. At the end of the session, the computer program randomly selected one of the four games to be payoff-relevant.

Player B's decision Player B decided under a veil of ignorance, *i.e.*, assuming that player A had chosen *In*. For each game, in treatment A (resp., treatment C) player B made four decisions corresponding to player A's (resp., player C's) four possible first-order beliefs on the frequency of *Right* choices. Therefore, in treatment A (resp., treatment C), player B could condition his decision to the possible first-order beliefs of player A (resp., player C). Given that player A had chosen *In* in the game randomly selected to be payoff-relevant, the computer program implemented the player B's decision corresponding to the actual first-order belief of player A (resp., player C) in treatment A (resp., treatment C). To facilitate decision making, these first-order beliefs were presented in a fixed increasing order (see an example of a decision screen in Appendix B). This method of eliciting decisions conditional on another player's first-order belief corresponds to the menu method introduced by Khalmetski et al. (2015). It allows the experimenter to artificially induce second-order beliefs.[10]

Second-order belief elicitation We also elicited, for each game, player A's second-order beliefs on the frequency of players A choosing *In* according to the players B and C in their group for this game. In other words, player A had to guess player B's and player C's first-order belief. We also elicited, for each game, player B's second-order belief on the frequency of B players choosing *Right* after *In* according to player A and player C. For each role, one belief out of the four elicited in the four games was randomly selected at the end of the session and paid €1 if accurate.

4.2 Individual Preferences

We elicited social preferences via the Social Value Orientation (SVO) slider measure (Murphy et al., 2011). This allows to classify participants in different categories: selfish (maximize own payoff), competitive (maximize the difference between own and other payoff), inequality averse (minimize the difference between own and other payoff), prosocial (maximize joint-payoff) and altruistic (maximize other payoff). The SVO measure consists in fifteen allocation decisions between a decision-maker and a passive player. All participants made fifteen choices as the decision maker. They were paid for two randomly selected periods: one as the decision maker, one as the passive player.

We also have non-incentivized measures of individual preferences. At the end of the session, subjects completed the Guilt and Shame Proneness (GASP) questionnaire (Cohen et al., 2011). We were particularly attentive to the Guilt-Negative-Behavior-Evaluation sub-scale that assesses one's proneness to feel bad about how one acted. Moreover, we elicited self-reported risk aversion, patience and guilt proneness through a single question each. Risk aversion and patience are measured by the following questions, respectively: "Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?" (Dohmen et al., 2011), and "Are you generally an impatient person, or someone who always shows great patience?" (Vischer et al., 2013). We adapted Moulton et al. (1966) to phrase the question on guilt proneness in a similar manner as for risk aversion and patience: "Are you generally a person who easily feel guilty or is it difficult to make you feel guilty?". Subjects rated how "easy" it is to make them feel guilty on a scale from 0 to 10, *i.e.*, with the same rating scale used to answer the two questions on

[10]The use of the menu method is frequent in the experimental literature on guilt aversion (Khalmetski et al., 2015; Hauge, 2016; Balafoutas and Fornwagner, 2017; Bellemare et al., 2017; Dhami et al., 2017; Bellemare 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 A players' (C players') true beliefs to B players. However, it requires choosing between two evils: if A players (C players) 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 dierences that are hidden at the aggregate level (Khalmetski et al., 2015).

how "willing to take risk" and how "patient" they are. Finally, we collected standard socio-demographic characteristics, including gender, age, professional status, and number of past participations in economic experiments.

4.3 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, using the software Hroot (Bock et al., 2014). 288 subjects participated in a total of 17 sessions. 57% were female and the average age was 22 years old (S.D. = 4.77). Each session lasted about 75 minutes. Game payoffs in the four games of Figures 3-6 were in Experimental Currency Unit (ECU) where 10 ECU = €1. The average earnings were €17 (S.D. = 5.91), including payment for correct beliefs and €5 of show-up fee.

At their arrival in the lab, subjects were randomly assigned to a cubicle after drawing a tag in an opaque bag. The instructions (see Appendix B) were distributed before each part of the experiment. The first part consisted in the description of the games as well as the completion of a comprehension questionnaire. The experimenter made sure that all subjects had completed correctly the comprehension questionnaire before moving on to the second part. At the beginning of the second part, all subjects were informed of their role. Then, we elicited the players' first-order beliefs and A players' decisions. The third part corresponded to B players' decisions. Meanwhile, A and C players could play sudoku. In the fourth part, we elicited the A and B players' second-order beliefs and let C players play sudoku. In the final part, we elicited the players' social preferences through the SVO approach. Then, the players received a feedback on their payoff and the decisions that were payoff-relevant. Finally, they completed a non-incentivized questionnaire in order for us to collect further measures of individual preferences and socio-democraphic characteristics.

5 Results

In the following, we consider first, behavior of A-subjects, *i.e.*, subjects in the role of player A (Section 5.1) and next, behavior of B-subjects (Section 5.2). Additional results on beliefs about A-subjects and B-subjects' behavior can be found, respectively, in Section A.1 and Section A.2 of the Appendix. 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.

5.1 A-subjects' behavior

Table 2 reports the results from Logit regressions on the probability that A-subjects choose *In* in each of the four games in Figures 3-6. For each game, there are two specifications. In the first specification, we regress A-subjects' *In* choice on α_{AB}, *i.e.*, their first-order belief on the likelihood of *Right* choices (to test H.A.1), and on their SVO angle (to test H.A.2). In line with our assumption of lexicographic altruistic preferences for A-subjects, we consider the SVO angle as a proxy of their altruism toward the player who could be more advantaged in terms of material payoff by that choice, *i.e.*, B-subjects in the Investment game (ϕ_{AB}), and C-subjects in the three other games (ϕ_{AC}). We control for the treatment A-subjects belonged to and the order in which the four games were played. In the second specification, we add personality (self-reported risk aversion and patience) and demographic controls (age, gender, number of previous experiments attended, and economics major).

	Investment		Rev-Investment		Donation		Exploitation	
α_{AB}	2.555***	2.366**	-0.314	-0.713	3.888***	4.841***	-0.586	-1.093
	(0.905)	(1.032)	(0.723)	(0.814)	(1.098)	(1.435)	(0.792)	(0.900)
SVO angle	0.059***	0.053***	0.028	0.029	0.068***	0.073**	-0.023	-0.027
	(0.018)	(0.019)	(0.019)	(0.020)	(0.026)	(0.030)	(0.018)	(0.020)
Treatment A	0.832*	1.176**	0.476	0.396	-0.239	-0.033	-0.287	-0.127
	(0.488)	(0.579)	(0.481)	(0.533)	(0.656)	(0.732)	(0.493)	(0.548)
Order	0.057	0.215	-0.244	-0.237	0.056	0.254	0.485*	0.495
	(0.268)	(0.301)	(0.197)	(0.220)	(0.256)	(0.288)	(0.270)	(0.318)
Personality	No	Yes	No	Yes	No	Yes	No	Yes
Demographics	No	Yes	No	Yes	No	Yes	No	Yes
Constant	-2.451***	-5.263	0.850	1.299	-4.421***	-8.829***	0.601	2.204
	(0.916)	(3.420)	(0.624)	(1.843)	(1.149)	(3.323)	(0.845)	(2.236)
Observations	96	96	96	96	96	96	96	96

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. "Order" takes value 1 when the game was presented first, etc. "Personality" controls correspond to the subjects' self-reported risk aversion and self-reported-reported patience. "Demographics" controls correspond to the subjects' age, gender, number of previous experiments attended, and economics major.

Table 2: Logit on the likelihood that A-subjects choose *In* across games

As for first-order beliefs (H.A.1), in the Investment and the Donation game, the more A-subjects believed that *Right* was likely to be chosen by B-subjects, the more they chose *In*, with regression coefficients being highly significant regardless of the specification. Conversely, in the Reversed-Investment and the Exploitation game, first-order beliefs do not significantly influence the frequency of *In* choices. This might be due to the majority of A-subjects being selfish, as it is usually found for trustees in comparable experimental studies of trust games (see, e.g., Attanasi et al., 2013, 2019a). In fact, Figure 7 shows that in both the Reversed-Investment and Exploitation game, the predicted behavior of selfish A-subjects is *In* regardless of their α_{AB}. This belief-independent behavior by the bulk of selfish A-subjects prevents detecting the predicted positive correlation for non-selfish ones. With this, we conclude that H.A.1 is only verified for the Investment and Donation game, as it is formally stated in Result A.1.

Result A. 1. *As predicted, the frequency of* In *choices by A-subjects increases in their first-order belief about B-subjects choosing* Right *in the Investment and the Donation game. Conversely, in the Reversed-Investment and the Exploitation game, the frequency of* In *choices is not significantly influenced by A-subjects' first-order beliefs.*

Table 2 also allows to examine the effect of altruism (H.A.2) on the frequency of *In* choices. In the Investment and in the Donation game, it shows that the wider A-subject's SVO angle, the higher the likelihood to choose *In*, with this effect being highly significant regardless of the specification. In the Reversed-Investment and in the Exploitation game, it instead shows no significant effect of A-subject's SVO angle over the likelihood to choose *In*. As for Result A. 1, this absence of significance might be due to the over-representation of selfish A-subjects in our pool, who are predicted to choose *In* regardless of α_{AB}. However, recall that H.A.2 is meant to show that the motivation behind the *In* choice of player A is different in different games. Therefore, we highlight that finding a significant positive correlation between altruism sensitivity and trust in the two games where this was predicted and no correlation in the two games where a negative relation was predicted, provides a partial though solid support for H.A.2. This is formally stated in Result A.2.

Result A. 2. *In the Investment and the Donation game, where a positive choice-type correlation is predicted, we confirm a significant positive correlation between the frequency of* In *choices by A-subjects and their sensitivity to altruism measured through the SVO angle. In the Reversed-Investment and the Exploitation game, where a negative choice-type correlation is predicted, we find no significant correlation between the two aforementioned measures.*

As for controls over H.A.1-H.A.2, Table 2 shows that neither the treatment nor the order of games' play had a statistically significant effect on A-subjects' choices, with two exceptions (in the Investment game and the Exploitation game, respectively). Result A. 1 and Result A. 2 are robust to the inclusion of personality and demographic controls.

Table 3 displays the proportion of A-subjects' *In* choices disentangled by the four possible first-order beliefs on the likelihood of *Right* choices. The table shows that across our 96 A-subjects, the majority (ranging from 45% to 65% depending on the game) thought that none of the three possible B-subjects they could be paired with would choose *Right* after *In*, and only a small fraction (ranging from 3% to 9%) thought that the three of them would choose *Right* after *In*.

% of *In* choices	Inv.	Rev-Inv.	Don.	Exp.
If $\alpha_{AB} = 0$	34.61%	69.76%	4.25%	80.95%
	(52)	(43)	(47)	(63)
If $\alpha_{AB} = 1/3$	48.00%	72.41%	17.85%	60.00%
	(25)	(29)	(28)	(20)
If $\alpha_{AB} = 2/3$	81.00 %	73.33%	64.70%	33.33%
	(16)	(15)	(17)	(6)
If $\alpha_{AB} = 1$	66.66%	66.66%	50.00%	100.00%
	(3)	(9)	(4)	(7)
All (96)	46.87%	70.83%	20.83%	75.00%

Sample size in parenthesis

Table 3: Proportion of *In* choices across games, by first-order belief

In the spirit of the predictions of Figure 7 on the altruism sensitivity ϕ_{Ah}, and of the separation among selfish, lowly-altruistic and highly-altruistic A-subjects on which H.A.3 and H.A.4 rely, we disentangle A-subjects uniformly into these three categories according to their SVO angle. More precisely, we define as "selfish" those A-subjects with SVO angle in the interval $(Min, Median - 15\%)$ of the empirical distribution, as "lowly-altruistic" those with SVO angle in $(Median - 15\%, Median + 15\%)$, and as "highly-altruistic" those with SVO angle in $(Median + 15\%, Max)$.

To test H.A.3, we consider choices by A-subjects who thought that *Left* was the most likely choice of B-subjects, *i.e.*, when A-subjects' $\alpha_{AB} \leq 1/3$. We also rely on the SVO-based separation among selfish, lowly-altruistic and highly-altruistic A-subjects introduced above. Focusing on selfish A-subjects, we find that the proportion of those who chose *In* is significantly lower in the Donation than in the Reversed-Investment game (2.70% vs 59.38%; proportion test, $p = 0.000$). Focusing on selfish and lowly-altruistic A-subjects, we find that the proportion of those who chose *In* is significantly lower in the Donation than in the Exploitation game (3.77% vs 80.77%; $p = 0.000$). Focusing on lowly-altruistic and highly-altruistic A-subjects, we find that the proportion of those who chose *In* is significantly lower in the Donation than in the Investment game (15.79% vs 48.89%; $p = 0.001$). These observations also hold if we release the constraints on the SVO angle of A-subjects, *i.e.*, if we consider the whole sample of A-subjects (proportion test, $p = 0.000$ for the three comparisons). For the Donation-Investment comparison, this highlights absence of unpredicted differences in the residual sub-samples of A-subjects, where the predicted choice is *Out* in both games (see Figure 7). For the other two pairwise comparisons, the latter result indirectly confirms, within our sample, a negligible fraction of highly-altruistic subjects (*i.e.*, those for whom the predicted choice is *In* in the Donation game and *Out* in the other two games). All this allows us to conclude that H.A.3 is verified, as it is stated formally stated in Result A.3.

Result A. 3. *As predicted, for A-subjects who thought that* Left *was the most likely action of B-subjects, the frequency of* In *choices in the Donation game was significantly lower than: (i) in the Reversed-Investment game for selfish types; (ii) in the Exploitation game for selfish and lowly-altruistic types; (iii) in the Investment game for all altruistic types.*

Finally, to test H.A.4, we consider choices by A-subjects who thought that *Right* was the most likely choice of B-subjects, *i.e.*, when A-subjects' $\alpha_{AB} > 2/3$. We continue to rely on the SVO-based separation among selfish, lowly-altruistic and highly-altruistic A-subjects introduced above. We find that the proportion of A-subjects who chose *In* in the Investment game is not significantly different than in the Reversed-Investment game regardless of the altruistic type (proportion test: $p = 1.000$ for selfish, $p = 0.898$ for lowly-altruistic, and $p = 0.167$ for highly-altruistic types). Furthermore, the proportion of selfish and lowly-altruistic A-subjects (considered together) who chose *In* is higher in the Investment than in the Donation game (60% vs 50%; $p = 0.671$), with the difference being significant for selfish A-subjects (92.31% vs 65%; $p = 0.074$). Finally, the proportion of highly-altruistic A-subjects who chose *In* is significantly higher in the Investment than in the Exploitation game (100% vs 50%, $p = 0.021$).

If we release the constraints on the SVO angle of A-subjects, *i.e.*, considering the whole sample of A-subjects, we still find that the proportion of A-subjects choosing *In* in the Investment game is not significantly different than in the Reversed-Investment game (78.94% vs. 70.83%, $p = 0.544$), and higher than in the Donation (61.90%, $p = 0.240$) and the Exploitation game (69.23%, $p = 0.533$).[11] This highlights absence of unpredicted differences in the residual sub-samples of A-subjects (*i.e.*, with $\phi_{Ah} \in (0.30, 3.33]$), where the predicted choice is *In* in all games (see Figure 7). All this allows us to conclude that H.A.4 is essentially verified, as it is stated formally stated in Result A.4.

Result A. 4. *As predicted, for A-subjects who thought that* Right *was the most likely action of B-subjects, the frequency of* In *choices in the Investment game is: (i) not significantly different than in the Reversed-Investment game; (ii) significantly higher than in the Exploitation game for highly-altruistic types; (iii) higher, although not significantly so, than in the Donation game for all other types.*

[11] The last difference in proportion is not significant due to the small number of A-subjects with $\alpha_{Ah} \geq 2/3$ in the Exploitation game (only 13/96, see Table 3).

5.2 B-subjects' behavior

Figure 8 shows B-subjects' pattern of choices for the four induced second-order beliefs $\beta_{Bj} \in \{0, 1/3, 2/3, 1\}$,[12] disentangled by the four games (Investment, Reversed-Investment, Donation and Exploitation) and the two treatments ($j = A$ and $j = C$).

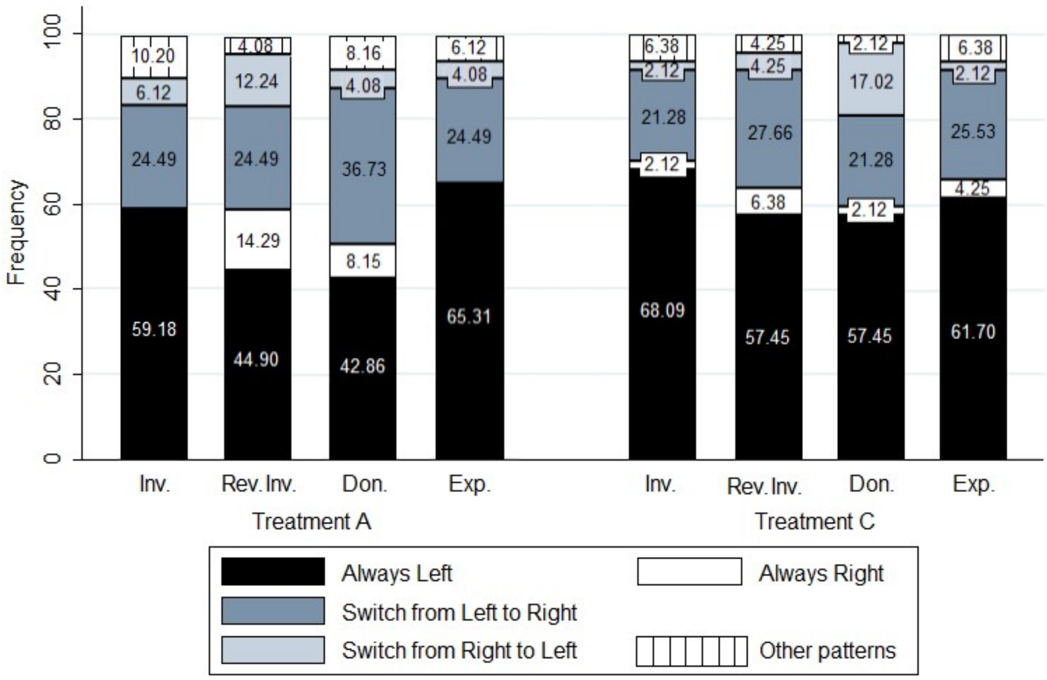


Figure 8: Distribution of B-subjects' pattern of choices across games and treatments

In each of the eight game-treatment combinations, we identified five patterns of choices: (i) always choosing *Left* regardless of the induced second-order beliefs, *i.e.*, choosing the payoff-maximizing (selfish) option, which represent on average 57% of B-subjects;[13] (ii) always choosing *Right* regardless of the induced second-order beliefs, *i.e.*, choosing the efficiency-maximizing option, on average 5%; (iii) switching from *Left* to *Right* as the induced second-order beliefs increase, *i.e.*, disclosing guilt aversion, on average 26%; (iv) switching from *Right* to *Left* as the induced second-order beliefs increase, on average 6%; and, (v) any other pattern of choices, on average 6%.

First of all, note that our model in Eqs. (4) to (6) captures on average 78% of all B-subjects' behavior, *i.e.*, patterns (i) to (iii) over (iv) and (v) in Figure 8. Furthermore, it explains 72% of the behavior of

[12] By "induced second-order beliefs" we denote the four possible first-order beliefs of A-subjects (resp., C-subjects) displayed on B-subjects' decision screens in the A treatment (resp., the C treatment).

[13] The fact that the fraction of selfish B-subjects detected in our four games is on average greater than 50% is not surprising. In fact, differently from the traditional trust game, B's trustworthiness (*Right* if *In*) brings him no additional money with respect to his initial endowment, since $\pi_B(R) = \pi_B(O)$. Therefore, in each of the four games, a B-subject choosing *Right* is purely driven by other-regarding preferences.

non-selfish subjects, *i.e.*, patterns (ii) and (iii) over (i)-(v) of B-subjects choosing *Right* for at least one induced second-order belief. Finally, and more importantly, guilt-averse behavior (ii) captures 60% of all non-selfish behavior (ii) to (v), thereby showing that guilt aversion is the prevailing social preference that is worth investigating in our four social-dilemma games.

We now test the five hypotheses elaborated in Section 3.3 about the behavior of B-subjects.

	Investment		Rev. Investment		Donation		Exploitation	
Induced β_{Bj}	2.406***	2.407***	1.759***	1.764***	1.694***	1.706***	3.040***	3.066***
	(0.555)	(0.556)	(0.478)	(0.479)	(0.442)	(0.443)	(0.638)	(0.638)
Stated β_{Bj}	3.504***	3.074***	4.102***	4.132***	3.523***	3.420***	3.595***	3.521***
	(0.934)	(0.903)	(0.990)	(0.987)	(0.960)	(0.894)	(0.955)	(0.925)
SVO angle	0.039*	0.026	0.091***	0.091***	0.038*	0.029	0.031	0.028
	(0.022)	(0.022)	(0.025)	(0.027)	(0.022)	(0.021)	(0.027)	(0.028)
Self-reported guilt	0.029	0.070	0.280**	0.316***	0.098	0.210**	0.068	0.190
	(0.101)	(0.104)	(0.112)	(0.116)	(0.093)	(0.093)	(0.122)	(0.126)
Treatment A	0.152	0.284	1.343**	1.161*	1.443***	1.591***	-0.463	-0.056
	(0.542)	(0.570)	(0.612)	(0.617)	(0.536)	(0.512)	(0.630)	(0.663)
Order	-0.174	-0.191	-0.096	-0.020	0.058	-0.124	-0.808**	-0.826**
	(0.296)	(0.307)	(0.231)	(0.236)	(0.192)	(0.192)	(0.334)	(0.327)
Personality	No	Yes	No	Yes	No	Yes	No	Yes
Demographics	No	Yes	No	Yes	No	Yes	No	Yes
Constant	-5.268***	-7.095**	-7.724***	-8.619***	-5.699***	-9.964***	-4.259***	-11.147***
	(1.460)	(2.976)	(1.391)	(2.831)	(1.115)	(2.412)	(1.489)	(3.702)
Observation	384	384	384	384	384	384	384	384
Subjects	96	96	96	96	96	96	96	96

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. "Induced β_{Bj}" corresponds to the four β_{Bj} presented to B-subjects when choosing *Right* or *Left*. "Stated β_{Bj}" corresponds to B-subjects stated β_{Bj} in the belief elicitation stage. "Order" takes value 1 when the game was presented first, etc. "Personality" controls correspond to the subjects' self-reported risk aversion and self-reported-reported patience. "Demographics" controls correspond to the subjects' age, gender, number of previous experiments attended, economics major.

Table 4: Panel Logit on the likelihood that B-subjects choose *Right* across games (Random Effects)

Table 4 presents the results from panel Logit regressions on the probability to choose *Right*. For each game, there are two specifications. In the first specification, we regress B-subjects' choices on their induced and their stated second-order belief β_{Bj} separately, their self-reported altruism through their SVO angle, and their self-reported guilt proneness from the final questionnaire. We control for the treatment they belonged to and the order in which the specific game was played. In the second specification, we add personality (self-reported risk aversion and patience) and demographic controls (age, gender, number of previous experiments attended, economics major).

As predicted by H.B.1., for all games, and for both specifications in Table 4, the higher the supposed first-order belief of their matched A-subject/C-subject, *i.e.*, B-subjects' induced second-order belief respectively in treatment A and C, the more likely B-subjects are to choose *Right*, at the 1% level of significance. The same holds for stated second-order beliefs: the likelihood of choosing *Right* significantly increases in B-subjects' stated second-order beliefs, again at the 1% level of significance and regardless of the specification in Table 4. Therefore, we conclude that H.B.1 is verified under each treatment and measurement of second-order beliefs, as it is stated formally stated in Result B.1 below.

Result B. 1. *The likelihood of B-subjects choosing Right significantly increases in their second-order beliefs about Right in each of the four games. All this holds independently from the treatment and both by considering induced and stated second-order beliefs.*

Regarding B-subjects' altruism sensitivity, we find that a higher SVO angle significantly increases the likelihood to choose *Right* in the Reversed Investment game and not in the Investment or Exploitation game, as predicted by H.B.2. Conversely, we do not detect a significant increase in the Donation game. However, if we look at Spearman correlations between the SVO angle and the number of *Right* choices for the four induced second-order beliefs that a B-subject makes, we find a significant positive correlation both in the Reversed-Investment and in the Donation game (respectively, $\rho_S = 0.42$, $p = 0.000$; $\rho_S = 0.30$, $p = 0.003$).

As for B-subjects' self-reported guilt sensitivity, although all regression coefficients are positive for each game and under each specification in Table 4, they have a significant impact on the likelihood to choose *Right* only in the Reversed-Investment and in the Donation game, and in the latter only if controlling for personality and demographic characteristics. Absence of significance in the other two games is not surprising, given that, differently from the external measure of altruism sensitivity (SVO angle), the external measure of guilt sensitivity was not incentivized (see Bellemare et al., 2019, for the difficulty of finding empirical relations between the concept of guilt aversion in economics and its characterization in psychological questionnaires). Therefore, this measure might also account for B-subjects' altruism, and the Reversed-Investment and the Donation game are the only two games where B's lexicographic altruism is triggered (see Eq. (5)). With this, we conclude that H.B.2 is verified in all games for altruism sensitivity and only in the Reversed-Investment and Donation game for guilt sensitivity, as it is stated formally in Result B.2 below.

Result B. 2. *The likelihood of B-subjects choosing Right increases in: (i) their altruism sensitivity measured through the SVO angle in the Reversed-Investment and the Donation game; (ii) their self-*

reported guilt sensitivity in the Reversed-Investment and the Donation game. All this holds independently from the treatment.

As for controls over H.B.1-H.B.2, Table 4 shows that treatment A had a positive effect on the likelihood that B-subjects choose *Right* in both the Reversed-Investment and Donation game. Note these are the only two games where the passive player C is ex-post vulnerable (see Table 1) and where B's altruism is activated (see Eq. (5)). The finding that B's other-regarding behavior toward C is boosted when we force him to focus on A's beliefs and that this only occurs in the two games where he can increase C's payoff can be seen as further support to the assumption of lexicographic altruism and to H.B.2. Also note that the order of game presentation has a negative effect on B's other-regarding behavior in the Exploitation game. Intuitively, being this game the only one where A's *In* choice is uncontroversially selfish, the later it is presented in the experiment the more B-subjects have faced less selfish *In* choices by A-subjects through induced second-order beliefs in the other games. Therefore, they show less trustworthiness in the Exploitation game. Results B.1 and B.2 are robust to the inclusion of personality and demographic controls.

To test H.B.3, we look at B-subjects' patterns of choices across games and treatments in Figure 8. More precisely, we focus on "Switch from *Left* to *Right*" patterns. According to a Kruskal-Wallis test, we find no significant difference in the proportion of guilt-averse B-subjects across games when we pool the treatments ($p = 0.504$) and within treatment C ($p = 0.632$). This is confirmed by Mann-Whitney tests of pairwise comparisons (lowest p-value with pooled treatments, $p = 0.133$; lowest p-value within treatment C, $p = 0.257$). Yet, in treatment A, there is a significant difference (at the 10% level) in the proportion of guilt-averse B-subjects driven by choices in the Donation game (Investment, Reversed-Investment and Exploitation: 24.49% *vs.* Donation: 36.73%; $p = 0.086$). With this, we can state that H.B.3 is essentially verified, as it is stated in Result B.3 below.

Result B. 3. *The proportion of guilt-averse B-subjects is not significantly different across the four games both in treatment C and by pooling both treatments. The same holds for all games but the Donation game in treatment A.*

Next, we compare the proportion of guilt-averse B-subjects across treatments within the same game. Again, we focus on "Switch from *Left* to *Right*" patterns in Figure 8, by comparing the two panels in the figure maintaining the game fixed. We find that the proportion of guilt-averse B-subjects is not different across treatments in the Investment, Reversed-Investment and Exploitation game (Investment: $p = 0.454$; Reversed-Investment: $p = 0.479$; Exploitation: $p = 0.813$), while in the Donation game, it is significantly greater in treatment A (36.73% *vs.* 21.28%, $p = 0.001$). Note that this significant difference confirms, once again, the importance of our extension of Battigalli and Dufwenberg (2007) to ex-ante vulnerability. In fact, the only significant difference that we find in the detected sensitivity to guilt is in favor of guilt triggered by ex-ante vulnerability (i.e., toward player A's belief that B will transfer her money to player C) than by ex-post vulnerability (i.e., toward player C's belief that B will transfer her the money sent by player A) in the Donation game. We recall that the former guilt-averse behavior is not predictable within the framework of Battigalli and Dufwenberg (2007), who only consider guilt triggered by ex-post vulnerability. This unpredicted result is in line with both Attanasi et al. (2019b), who find an effect of guilt aversion on the second-mover's behavior also when the payoff of the disappointed player is not affected. It is also in line with the findings of Bellemare et al. (2017), who detect more guilt among trustees than among dictators: in our three-player games, this translates into a higher guilt sensitivity of B-subjects toward a player signaling her intentions through her previous move (A, similar to a trustor in a trust game) than toward the passive player C (similar to a recipient in a dictator game). With this, we confirm H.B.4 for all games but the Donation game, as stated below in Result B.4.

Result B. 4. *The proportion of guilt-averse B-subjects is not significantly different across treatments within the Investment, Reversed-Investment and Exploitation game. In the Donation game, the proportion of guilt-averse B-subjects is significantly greater under ex-ante vulnerability (treatment A) than under ex-post vulnerability (treatment C) of the disappointed player.*

We now turn to H.B.5, which concerns the proportion of B-subjects always choosing *Left*, *i.e.*, the disclosed proportion of selfish B-subjects. Looking at the fraction of "Always *Left*" B-subjects in Figure 8, we find that, pooling the two treatments, the proportion of selfish B-subjects in the Investment game is higher than in the Reversed-Investment and Donation game (McNemar test; Investment *vs.* Reversed-Investment: $p = 0.007$; Investment *vs.* Donation: $p = 0.004$). Furthermore, this proportion is also significantly higher in the Exploitation game than in the Reversed-Investment and Donation game (McNemar test; Exploitation *vs.* Reversed-Investment: $p = 0.018$; Exploitation *vs.* Donation: $p = 0.009$). The higher proportion of selfish B-subjects in the Investment and Exploitation game is still detected disentangling by treatment, although the differences with respect to the other two games are significant in treatment A (McNemar test; highest $p = 0.032$ for Investment *vs.* Donation game) but not in treatment C (the only significant difference is found between Investment *vs.* Donation game, $p = 0.058$). Therefore, we conclude that H.B.5 is verified, as it is stated formally in Result B.5. below.

Result B. 5. *The proportion of selfish B-subjects is significantly higher in the Investment and Exploitation games than in the Reversed Investment and Donation games. This holds independently from the treatment, although significantly only in treatment A.*

Finally, following Bellemare et al. (2011) and Attanasi et al. (2019b), we define a structural econometric model to estimate player B's average guilt sensitivity γ_{Bjk} toward player j's beliefs about player k's payoff, with $j, k \in \{A, C\}$ in each of the eight game-treatment combinations. In line with our model in Eqs. (4) and (6), each B-subject chooses between strategies *Right* if *In* and *Left* if *In* for each of the four possible first-order beliefs of player j about *Right* if *In* (α_{jB}), in order to maximize his utility as defined by Eq. (10) (Random Utility Model). In this equation, λ is the noise parameter that we estimate and U_B follows our model in Eq. (4), for $\alpha_{jB} \in \{0, 1/3, 2/3, 1\}$, and $j \in \{D, R\}$:[14]

$$V_B(\gamma_{Bjk}, \lambda, z|In) \quad = \quad U_B(\gamma_{Bjk}, z|In) + \lambda \cdot \epsilon_B(z|In) \tag{10}$$

We used a conditional Logit model to estimate γ_{Bjk}, the sensitivity to $\max\{0, \mathbb{E}_j[\pi_k(z|In)] - \pi_k(z|In)\}$ in Eq. (6), while fixing to 1 the coefficient corresponding to B's material payoff $\pi_B(z|In)$. In Table 5, we report the structural estimates of average guilt sensitivity in each game-treatment combination, by only considering B-subjects exhibiting a behavior consistent with our model predictions, *i.e.*, choosing always *Left* or always *Right* regardless of the four α_{jB}, or switching from *Left* to *Right* as α_{jB} increases (see Figure 8). On average, those B-subjects represent 87.54% of experimental participants in this role.

	Treatment A				Treatment C				All
	Inv.	Rev-Inv.	Don.	Exp.	Inv.	Rev-Inv.	Don.	Exp.	
γ_{Bjk}	-0.43***	-0.45***	-0.50***	-0.39***	-0.34***	-0.39***	-0.36***	-0.36***	-0.39***
	(0.03)	(0.06)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)	(0.01)
Obs.	328	328	344	352	304	344	344	344	2688

Table 5: Structural estimates of guilt sensitivity for B-subjects disclosing behavior consistent with the model

Pooling all games and treatments, we find that, on average, B-subjects are willing to pay 0.39 ECU to avoid disappointing their co-player's expectations by 1 ECU. Confidence intervals of the estimated γ_{Bjk} under the eight different specifications always overlap with the confidence interval of our benchmark (All). It appears that there is no combination of game and treatment where the desire to avoid disappointing a co-player is greater or lower than the average. More generally, when comparing estimated γ_{Bjk} per

[14] Recall that, differently from Bellemare et al. (2011), in our theoretical model B can also be altruistic toward player C, with altruism sensitivity measured through the parameter ϕ_{BC} (see Eq. (5)). However, the second component of B's feeling of altruism F_{BC}, *i.e.*, C's material payoff π_C, is colinear with B's material payoff π_B (by construction of the four games). Therefore, we cannot estimate the three coefficients (ϕ_{BC}, γ_{Bjk}, and the coefficient corresponding to π_B) of our theoretical utility function while estimating the noise parameter of our random utility model in (Eq. (10)). Therefore, we renounce to estimate ϕ_{BC} in the two games where it is assumed to be non-null, *i.e.*, in the Reversed-Investment and in the Donation game. In the remaining two games, this is not an issue, since by construction $\phi_{BC} = 0$ (lexicographic altruism).

game and treatment, confidence intervals always overlap. Therefore, the structural estimates in Table 5 essentially confirm Results B.3 and B.4 above and provide further support to H.B.3 and H.B.4.

6 Conclusion

This paper contributes to the literature investigating the facilitating factors of guilt aversion in the lab. We have designed an experiment to test whether the structure of the game could facilitate or attenuate the expression of guilt aversion. In particular, in a series of four three-player trust games we identify different types of players' vulnerability as potential determinants of the propensity of the trustee to feel guilty towards the other two players. These games are the classical Investment game, the Reversed-Investment game, the Donation game, and the Exploitation game. To the best of our knowledge, guilt aversion was never tested in Reversed-Investment and the Exploitation game, and only recently assessed in the Donation game (Attanasi et al., 2019b). For each of the four games, we implement two treatments: one where the trustee's choice is elicited conditionally to the expectations of the trustor, and another where it is elicited conditionally to those of the passive player.

Results of our laboratory experiment show that the trustor's behavior is dependent on the game structure of the four strategic settings, disclosing different levels of trust according to ex-ante and/or ex-post vulnerability of her and/or of the passive player. These results are in line with the predictions of our theoretical model, where we assume that the trustor can feel altruism toward the trustee in the Investment game (where the passive player is just an external observer) and toward the passive player in the other three games, thereby trusting more the more she believes the trustee to be trustworthy. Despite a different interplay of altruism sensitivity and beliefs in the four different games, expectation of trustworthiness is the channel for the activation of the trustor's disappointment, which is the primitive of trustee's guilt aversion through his second-order beliefs.

Indeed, for the trustee we find a positive correlation between his trustworthiness and his second-order belief of trustworthiness in all four games regardless of the treatment, thereby confirming the relevance of guilt aversion under every manipulation of the other two players' type of vulnerability. Furthermore, and more importantly, once disentangled from game-dependent purely altruistic behavior, the proportion of guilt-averse trustees is detected not to be significantly different either across the four games or across the two treatments. In particular, the trustees show a guilt-averse behavior also toward the beliefs of players who are not vulnerable at all.

The latter result leads us to conclude that guilt aversion is insensitive to disappointing payoff-vulnerable players or that vulnerability plays a secondary effect with respect to guilt driven by disappointing another player's beliefs. Rather, in line with the intuition of Attanasi et al. (2016) which postulate that guilt is role-dependent in the trust game, we find that the role effect is so pervasive that the trustee's guilt sensitivity is insensitive to manipulations of the trustor's payoffs and intentions.

Finally, the role of players' beliefs behind guilt aversion seems to be so crucial that guilt is activated even when the beliefs of the disappointed player do not concern her material payoff but the payoff of a third player, which is the key contribution of our theoretical extension of Battigalli and Dufwenberg (2007) provided by our portable model in Eq. (4)-(6). With this, we strengthen the view that specific belief-dependent feelings can be systematically ascribed to specific game-role combinations in asymmetric games.

References

Andrighetto, G., Grieco, D., and Tummolini, L. (2015). Perceived legitimacy of normative expectations motivates compliance with social norms when nobody is watching. *Frontiers in Psychology*, 6:1413.

Attanasi, G., Battigalli, P., and Manzoni, E. (2016). Incomplete-information models of guilt aversion in the trust game. *Management Science*, 62(3):648–667.

Attanasi, G., Battigalli, P., Manzoni, E., and Nagel, R. (2019a). Belief-dependent preferences and reputation: Experimental analysis of a repeated trust game. *Journal of Economic Behavior & Organization*, 167:341–360.

Attanasi, G., Battigalli, P., and Nagel, R. (2013). Disclosure of belief-dependent preferences in a trust game. Technical report, No. 506, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.

Attanasi, G., Rimbaud, C., and Villeval, M. C. (2019b). Embezzlement and guilt aversion. *Journal of Economic Behavior & Organization*, 167:409–429.

Balafoutas, L. (2011). Public beliefs and corruption in a repeated psychological game. *Journal of Economic Behavior & Organization*, 78(1-2):51–59.

Balafoutas, L. and Fornwagner, H. (2017). The limits of guilt. *Journal of the Economic Science Association*, 3(2):137–148.

Battigalli, P., Charness, G., and Dufwenberg, M. (2013). Deception: The role of guilt. *Journal of Economic Behavior & Organization*, 93:227–232.

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

Battigalli, P. and Dufwenberg, M. (2019). Psychological game theory. Technical report, No. 646 IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.

Baumeister, R. F., Stillwell, A. M., and Heatherton, T. F. (1994). Guilt: an interpersonal approach. *Psychological Bulletin*, 115(2):243.

Beck, A., Kerschbamer, R., Qiu, J., and Sutter, M. (2013). Shaping beliefs in experimental markets for expert services: Guilt aversion and the impact of promises and money-burning options. *Games and Economic Behavior*, 81:145–164.

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

Bellemare, C., Sebald, A., and Suetens, S. (2018). Heterogeneous guilt sensitivities and incentive effects. *Experimental Economics*, 21(2):316–336.

Bellemare, C., Sebald, A., and Suetens, S. (2019). Guilt aversion in economics and psychology. *Journal of Economic Psychology*, 73:52–59.

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

Bock, O., Baetge, I., and Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, 71:117–120.

Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American economic review*, 90(1):166–193.

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

Buskens, V. and Raub, W. (2013). Rational choice research on social dilemmas: embeddedness effects on trust. *Handbook of Rational Choice Social Research*, pages 113–150.

Cartwright, E. (2018). A survey of belief-based guilt aversion in trust and dictator games. *Journal of Economic Behavior & Organization*.

Cartwright, E. (2019). Guilt aversion and reciprocity in the performance-enhancing drug game. *Journal of Sports Economics*, 20(4):535–555.

Charness, G. and Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, 74(6):1579–1601.

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

Chen, J. and Houser, D. (2018). Broken promises and hidden partnerships: An experiment. *Journal of Economic Behavior & Organization*.

Cohen, T. R., Wolf, S. T., Panter, A. T., and 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.

Cox, J. C., Kerschbamer, R., and Neururer, D. (2016). What is trustworthiness and what drives it? *Games and Economic Behavior*, 98:197–218.

Danilov, A., Khalmetski, K., and Sliwka, D. (2019). Descriptive norms and guilt aversion.

Dhami, S., Wei, M., and al Nowaihi, A. (2017). Public goods games and psychological utility: Theory and evidence. *Journal of Economic Behavior & Organization*.

Di Bartolomeo, G., Dufwenberg, M., and Papa, S. (2019). The sound of silence: A license to be selfish. *Economics Letters*, 182:68–70.

Di Bartolomeo, G., Dufwenberg, M., Papa, S., and Passarelli, F. (2018). Promises, expectations & causation. *Games and Economic Behavior*.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Dufwenberg, M. and Patel, A. (2019). Introduction to special issue on psychological game theory. *Journal of Economic Behavior & Organization*.

Ederer, F. and Stremitzer, A. (2017). Promises and expectations. *Games and Economic Behavior*, 106:161–178.

Engler, Y., Kerschbamer, R., and Page, L. (2018). Why did he do that? using counterfactuals to study the effect of intentions in extensive form games. *Experimental Economics*, 21(1):1–26.

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

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

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

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

Kawagoe, T. and Narita, Y. (2014). Guilt aversion revisited: An experimental test of a new model. *Journal of Economic Behavior & Organization*, 102:1–9.

Khalmetski, K. (2016). Testing guilt aversion with an exogenous shift in beliefs. *Games and Economic Behavior*, 97:110–119.

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

Morell, A. (2014). The short arm of guilt: Does it only hit who is close? Technical report, Preprints of the Max Planck Institute for Research on Collective Goods.

Moulton, R. W., Burnstein, E., Liberty Jr, P. G., and 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.

Murphy, R. O., Ackermann, K. A., and Handgraaf, M. (2011). Measuring social value orientation. *Judgment and Decision Making*, 6(8):771–781.

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

Peeters, R., Vorsatz, M., and Walzl, M. (2015). Beliefs and truth-telling: A laboratory experiment. *Journal of Economic Behavior & Organization*, 113:1–12.

Pelligra, V., Reggiani, T., and Zizzo, D. J. (2016). Responding to (un) reasonable requests. *Available at SSRN 2835557*.

Reuben, E., Sapienza, P., and Zingales, L. (2009). Is mistrust self-fulfilling? *Economics Letters*, 104(2):89–91.

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., Sunde, U., and Wagner, G. G. (2013). Validating an ultra-short survey measure of patience. *Economics Letters*, 120(2):142–145.

Appendix

A Additional Results

A.1 Belief on A-subjects' behavior across games

Table 6 displays the first-order belief of B-subjects and C-subjects about the proportion of A-subjects choosing *In*. There is no difference between the average first-order belief of B-subjects and C-subjects in either of the games (Mann-Whitney rank sum tests, smallest $p = 0.333$).

	Inv.	Rev-Inv.	Don.	Exp.
B-subjects' FOB(*In*)	43.41%	65.69%	18.62%	81.79%
	(30.96)	(34.23)	(29.61)	(29.89)
C-subjects' FOB(*In*)	45.85%	62.16%	19.63%	80.76%
	(32.12)	(31.83)	(28.67)	(31.34)

Standard errors in parenthesis (in % points)

Table 6: B-subjects' and C-subjects' first-order beliefs on the proportion of A-subjects choosing *In*

Table 7 shows that, in the Donation game and the Investment game, there is no significant difference between the average of B-subjects' and C-subjects' beliefs about the proportion of A-subjects choosing *In* and the true proportion; whereas, in the Reversed-Investment game and the Exploitation game, B-subjects and C-subjects significantly overestimated, and respectively underestimated, the frequency of *In* choices (t-tests, $p = 0.004$, and respectively, $p = 0.004$).

	Inv.	Rev-Inv.	Don.	Exp.
Significant difference between FOB and choices (t-tests)	No	Yes***	No	Yes***
Frequency of underestimation of *In* choices	55.73%	66.15%	63.02%	33.33%
Frequency of overestimation of *In* choices	44.27%	33.85%	36.98%	66.66%
Total	100%	100%	100%	100%

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

Table 7: Differences between B-subjects' and C-subjects' first-order beliefs and A-subjects' choice to go *In*

A.2 Beliefs on B-subjects' behavior across games

Table 8 displays the first-order belief of A-subjects and C-subjects about the proportion of B-subjects choosing *Left*. In the Donation game, A-subjects believe that there will be significantly less *Left* choices than C-subjects do (Mann-Whitney rank sum test, $p = 0.086$). For the other games, there is no difference between the average first-order belied of A-subjects and C-subjects (Mann-Whitney rank sum tests, smallest $p = 0.288$).

	Inv.	Rev-Inv.	Don.	Exp.
A-subjects' FOB(*Left*)	77.28%	70.34%	75.45%	81.70%
	(28.53)	(32.84)	(29.41)	(30.00)
C-subjects' FOB(*Left*)	81.76%	75.21%	81.42%	82.43%
	(25.79)	(28.81)	(26.17)	(26.30)

Standard errors in parenthesis (in % points).

Table 8: A-subjects' and C-subjects' first-order beliefs on the proportion of B-subjects choosing *Left* across games

Table 9 shows that there is a significant difference between the average of A-subjects' and C-subjects' beliefs about the proportion of B-subjects choosing *Left* and the true proportion for each game (t-tests, each $p = 0.000$).[15]

	Inv.	Rev-Inv.	Don.	Exp.
Significant difference between FOB and choices (t-tests)	Yes***	Yes***	Yes***	Yes***
Frequency of underestimation of *Left* choices	16.67%	23.44%	19.27%	35.94%
Frequency of overestimation of *Left* choices	83.33%	76.57%	80.73%	64.06%
Total	100%	100%	100%	100%

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

Table 9: Differences between A-subjects' and C-subjects' first-order beliefs and B-subjects' choice to go *Left*

[15] We compare A-subjects' and C-subjects' first order belief to the proportion of B-subjects choosing *Left* for the true first-order belief of player A (respectively, player C) in treatment A (respectively, treatment C).

A.3 B-subjects' patterns of choices within-individuals

We propose to classify individuals among three categories: (i) consistent, *i.e.*, B-subjects who always had the same pattern of choices across the four games: 58.08% of B-subjects; (ii) nearly consistent, i.e. B-subjects who had the same pattern of choices in three games: 20.83%; and, (iii) inconsistent, i.e. B-subjects who had the same pattern of choices in two games at most: 27.08%. Table 10 lists the pattern of choices of the 26 participants who were inconsistent across games.

	Id	Inv.	Rev-Inv.	Don.	Exp.
Treatment A	1	Always *Left*	*Right* to *Left*	*Left* to *Right*	Always *Right*
	2	Other	*Left* to *Right*	Always *Left*	Always *Left*
	3	*Left* to *Right*	Always *Right*	Always *Right*	Always *Left*
	4	Always *Left*	Always *Right*	Always *Right*	Always *Left*
	5	Other	Always *Right*	*Left* to *Right*	Always *Left*
	6	Other	Always *Right*	Always *Right*	*Left* to *Right*
	7	Always *Left*	Always *Right*	*Left* to *Right*	Always *Left*
	8	Always *Left*	*Right* to *Left*	Other	Always *Left*
	9	Always *Left*	Always *Right*	Always *Right*	Always *Left*
	10	*Left* to *Right*	Always *Left*	*Left* to *Right*	Always *Left*
	11	Always *Left*	Always *Left*	*Left* to *Right*	*Left* to *Right*
	12	*Left* to *Right*	*Left* to *Right*	Always *Left*	Always *Left*
	13	Always *Left*	Other	*Left* to *Right*	Always *Left*
	14	*Left* to *Right*	*Left* to *Right*	Always *Left*	Always *Left*
Treatment C	15	Always *Right*	Always *Left*	Always *Left*	Always *Right*
	16	Always *Left*	*Right* to *Left*	*Right* to *Left*	Always *Left*
	17	Other	*Left* to *Right*	*Left* to *Right*	Other
	18	Always *Left*	*Left* to *Right*	*Left* to *Right*	Always *Left*
	19	Always *Left*	Always *Right*	Always *Right*	Other
	20	*Left* to *Right*	Other	*Left* to *Right*	Always *Left*
	21	*Left* to *Right*	Always *Left*	*Right* to *Left*	*Left* to *Right*
	22	Always *Left*	*Left* to *Right*	Always *Left*	*Left* to *Right*
	23	Always *Left*	Other	*Right* to *Left*	Always *Left*
	24	Other	Always *Right*	*Right* to *Left*	Other
	25	Always *Left*	*Left* to *Right*	Always *Left*	*Left* to *Right*
	26	Other	*Left* to *Right*	*Right* to *Left*	Always *Left*

Table 10: B-subjects showing inconsistent patterns, *i.e.*, with at most two similar belief-dependent patterns across games

Figure 9 focuses on the left (respectively, right) on B-subjects that were classified as guilt-averse (respectively, selfish) in at least one game. For both type of preferences, B-subjects who are consistent in at least three games constitute the majority of our observation: 56.10% for "guilt-averse" subjects and 68,18% for "selfish" subjects.

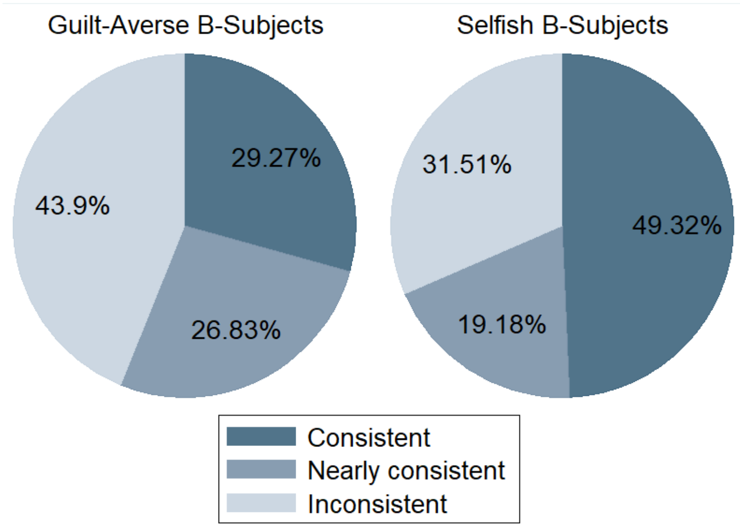


Figure 9: Distribution of B-subjects consistency of behavior

B Instructions

We 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 decisions of other participants. At the end of the session, you will receive your earnings in cash in a separate room to preserve the confidentiality of your earnings. The earnings you will receive will include:

- your earnings from today's experimental session
- a €5 fee for showing-up on time to the experimental session

During the session, the currency used is sometimes ECU (Experimental Currency Units).

Please turn off your phone. Communication with other participants is prohibited during the entire duration of the session. If you have questions during the session, raise your hand or press the red button on the side of your desk and we will come to answer in private.

OVERVIEW OF THE SESSION

In this session, there are two parts. Both parts are completely independent.

In each part, one or more of your decisions will be randomly selected by the computer. At the end of the session, you will be informed of your decisions, the decisions of other participants (if these affect your winnings) and their impact on your earnings.

At the end of the session you will have to answer a final questionnaire.

FIRST PART: OVERVIEW

In this part, the conversion rate is as follows: 10 ECU = €1.

The roles

At the beginning of the first part, the computer program randomly assigns a role to each participant. You can be either a participant A, a participant B or a participant C. Your role is indicated on your computer screen when the first starts and it remains the same during the first part.

Then, the computer program randomly forms groups of three participants, with one participant of each role in each group. The computer program forms a new group for each situation (which we will describe below) so your group changes during the first part. You will never know the identity of other members of your group and they will never be informed yours.

The decisions

Each participant receives an initial endowment. First, the participant A must make a decision. He can send 25 ECU to the participant B or not. The 25 ECU sent to Participant B come from the endowment of either the participant A or the participant C, depending on the situation.

Then, if the participant B has received 25 ECU, he must make a decision. He decides how to distribute these 25 ECU between another participant (A or C, depending on the situation) and himself. The ECU that the participant B transfers to another participant (A or C, depending on the situation) are multiplied by two while the ECU that the participant B keeps for himself are not multiplied by two.

The situations

There are four different situations: the situation "North", the situation "West", the situation "East" and the "South" situation (the name of each situation was given arbitrarily). You will make decisions in each of these four situations.

In the North situation, the participant A decides whether or not to send 25 ECU from his initial endowment to the participant B. If the participant B receives these 25 ECU, he decides how to distribute these 25 ECU between the participant C and himself.

In the situation, the participant A decides whether or not to send 25 ECU of his initial endowment to the participant B. If the participant B receives these 25 ECU, he decides how to distribute these 25 ECU between the participant A and himself.

In the East situation, the participant A decides whether or not to send 25 ECU from the initial endowment of the participant C to the participant B. If the participant B receives these 25 ECU, he decides how to distribute these 25 ECU between the participant C and himself.

In the South situation, the participant A decides whether or not to send 25 ECU from the initial endowment of the participant C to the participant B. If the participant B receives these 25 ECU, he decides how to distribute these 25 ECU between the participant A and himself.

Now we will detail the roles, decisions and situations of the first part.

FIRST PART: ROLES, DECISIONS, SITUATIONS

The participant A

The participant A receives an initial endowment of 170 ECU.

He decides whether or not to send 25 ECU from either his endowment or the participant's C endowment to the participant B.

In the North situation, the participant A has the choice between:

- sending 25 ECU from his initial endowment to the participant B
- sending 0 ECU from his initial endowment to the participant B

In the West situation, the participant A has the choice between:

- sending 25 ECU from his initial endowment to the participant B
- sending 0 ECU from his initial endowment to the participant B

In the East situation, the participant A has the choice between:

- sending 25 ECU from the participant's C initial endowment to the participant B
- sending 0 ECU from the participant' C initial endowment to the participant B

In the South situation, the participant A has the choice between:

- sending 25 ECU from the participant's C initial endowment to the participant B
- sending 0 ECU from the participant' C initial endowment to the participant B

The participant B

The participant B receives an initial endowment of 100 ECU.

If the participant A has sent 25 ECU to the participant B, the participant B makes a decision.

The participant B decides how to distribute the 25 ECU between another participant (A or C, depending on the situation) and himself. The ECU that participant B transfer to another participant (A or C, depending on the situation) are doubled, while the ECU that the participant B keeps for him are not doubled.

In the North situation, the participant B has the choice between:

- transferring the 25 ECU to the participant C - the participant C receives 50 ECU
- transferring 5 ECU to the participant C - the participant C receives 10 ECU - and keeping 20 ECU for himself - the participant B keeps 20 ECU.

In the West situation, the participant B has the choice between:

- transferring the 25 ECU to the participant A - the participant A receives 50 ECU
- transferring 5 ECU to the participant A - the participant A receives 10 ECU - and keeping 20 ECU for himself - the participant B keeps 20 ECU.

In the East situation, the participant B has the choice between:

- transferring the 25 ECU to the participant C - the participant C receives 50 ECU
- transferring 5 ECU to the participant C - the participant C receives 10 ECU - and keeping 20 ECU for himself - the participant B keeps 20 ECU.

In the South situation, the participant B has the choice between:

- transferring the 25 ECU to the participant A - the participant A receives 50 ECU
- transferring 5 ECU to the participant A - the participant A receives 10 ECU - and keeping 20 ECU for himself - the participant B keeps 20 ECU.

If the participant A has not sent 25 ECU to the participant B, the participant B does not make any decision.

The participant C

The participant C receives an initial endowment of 30 ECU.

Irrespective of the situation, the participant C does not make any decision.

FIRST PART: STAGES

The first part of this session consists of four stages:

- Stage 1: All participants answer to some questions.
- Stage 2: The participant A makes his decisions in the four situations.
- Stage 3: The participant B makes his decisions in the four situations.
- Stage 4: The participant A and the participant B answer to some questions.

FIRST PART: COMPREHENSION QUESTIONNAIRE

Please complete the comprehension questionnaire that we will distribute you. If you have any difficulty answering the questionnaire or when you have completed the questionnaire, raise your hand or press the red button on the side of your desk. We will answer your questions in private.

——————————— End of the first set of instructions ———————————

STAGE 1

In this step, all participants answer some questions.

If you are a participant B or a participant C, you must answer the following question: "Out of 3 participants A randomly selected in today's session, how many of these participants A will send 25 ECU to the participant B?".

You must answer this question for each situation: North, West, East and South.

If you are a participant A or a participant C, you must answer the following question: "Out of 3 participants B randomly selected in today's session, if the participant A has sent 25 ECU, how many of these participants B will transfer the 25 ECU to another participant?".

You must answer this question for each situation: North, West, East and South.

How do the answers to these questions affect your earnings?

At the end of the session, for each role, one of the questions you answered during this stage will be randomly selected by the computer program. If your answer to this question is correct, you earn €1.

Example: Suppose you are a participant C and the randomly selected question is: "In the West situation, Out of 3 participants A randomly selected in today's session, how many of these participants A will send 25 ECU to the participant B?". The computer program randomly select 3 participants A among the participants A in this session. If, in the West situation, x participant(s) A among the 3 participants A randomly selected hashave decided to send 25 ECU to the participant B, then, your answer is correct if you answered "x".

STAGE 2

In this stage, participant A makes his decisions.

If you are a participant B or a participant C, you do not make any decision in this step.

If you are a participant A, you decides whether or not to send 25 ECU to the participant B.

You must make this decision in each situation: North, West, East and South.

Which decision of the participant A determines the earnings of the group members?

At the end of the session, the computer program randomly selects the situation North, West, East or South. The decision taken in the randomly selected situation determines the earnings of the group members.

At the end of the session, all group members are informed of the participant's A randomly selected decision.

I you have any questions raise your hand or press the red button on the side of your desk. We will answer your questions in private.

—————————— End of the second set of instructions ——————————

STAGE 3

In this stage, the participant B makes his decisions.

If you are a participant A or a participant C, you do not make any decision in this stage.

If you are a participant B, you decide how to distribute the 25 ECU you received between another participant (A or C) and yourself.

You must make this decision in each situation: North, West, East and South. Furthermore, in each situation, you have to make that decision depending on the prediction of the participant *A/C*.[16] To better understand, look at the screen example below. There are two pieces of information which appear in bold on the screen: a piece of information on the situation and a piece of information on the prediction of the participant C.

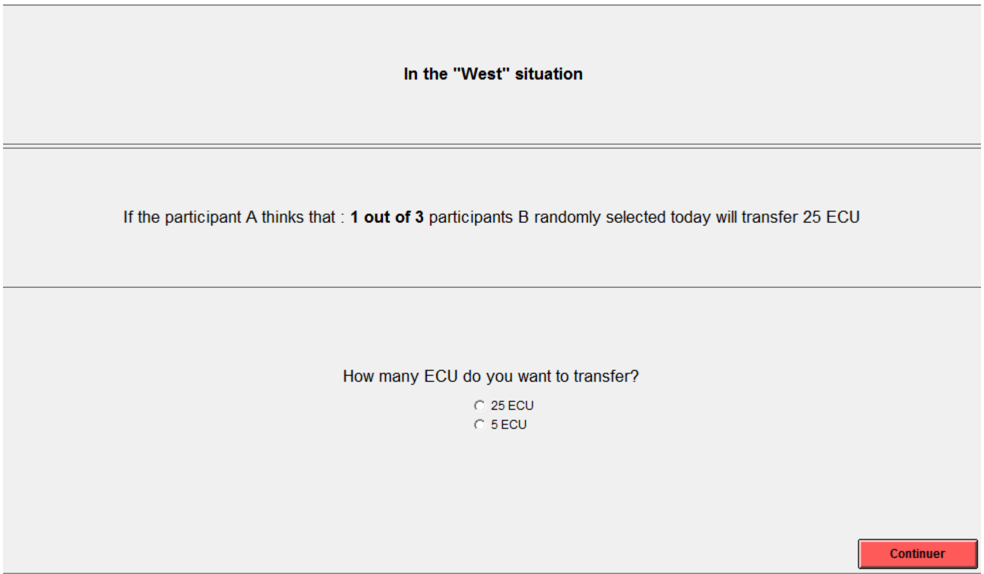


Figure 10: Screenshot in Treatment A

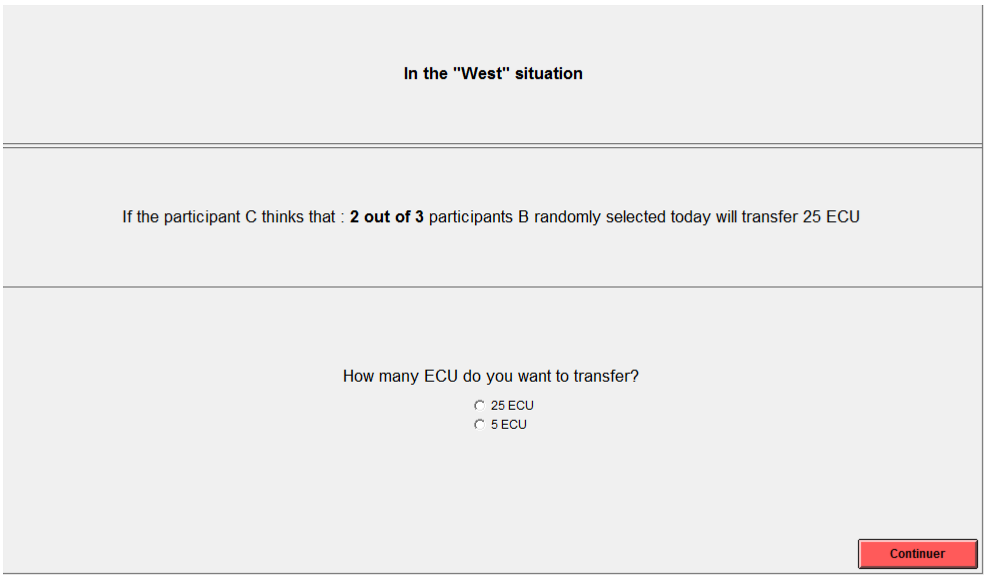


Figure 11: Screenshot in Treatment C

Information on the situation

You decide how to distribute the 25 ECU for each situation.

[16] Text between *... / ...* represents the two versions of the instructions. The first version corresponds to the Treatment A and the second version corresponds to the Treatment C.

Example: In the screen above, you make your decision in the West situation.

Information on the prediction of the participant *A/C*

Remember that in the stage 1, the participant *A/C* answered the following question for each situation: "Out of 3 participants B randomly selected in today's session, if the participant A has sent 25 ECU, how many of these participants B will transfer the 25 ECU to another participant?". There were four possible predictions: 0, 1, 2 or 3. You decide how to distribute the 25 ECU for each of the possible predictions of the participant *A/C*.

Example: In the screen above, you make your decision in the West situation and when the participant *A/C* of your group thinks 2 out of 3 participants B randomly selected today will transfer 25 ECU to another participant.

In total, you must therefore make 16 decisions:

- Four decisions corresponding to the four possible predictions of the participant *A/C* in the North situation
- Four decisions corresponding to the four possible predictions of the participant *A/C* in the West situation
- Four decisions corresponding to the four possible predictions of the participant *A/C* in the East situation
- Four decisions corresponding to the four possible predictions of the participant *A/C* in the South situation

However, only one of these decisions is susceptible to determine the earnings of the group members.

Which decision determines the earnings of the group members?

If the participant A has decided to send 0 EMU to the participant B, no decision of the participant B determines the earnings of the group members.

If the participant A has decided to send 25 EMU to the participant B, a decision of the participant B determines the earnings of the group members.

At the end of the session, the computer program randomly selects the situation North, West, East or South. Of the 4 decisions made by participant B in the selected situation, the computer program then selects the decision that corresponds to the prediction that the participant *A/C* made in stage 1. This decision determines the earnings of the group members.

At the end of the session, all group members are informed of the participant's B randomly selected decision.

Example: Suppose that the computer program randomly selects the West situation. Suppose that, to the question "In the West situation, out of 3 participants B randomly selected in today's session, if the participant A has sent 25 ECU, how many of these B participants will transfer the 25 ECU to another participant?", the participant *A/C* answered "2". Then, the computer program selects the decision that the participant B made when his screen displayed "West situation" and "The participant *A/C* thinks that 2 out of 3 participants B randomly selected today will transfer 25 ECU to another participant " (see the example screen above). This decision determines the earnings of the group members.

I you have any questions raise your hand or press the red button on the side of your desk. We will answer your questions in private.

——————— End of the third set of instructions ———————

STAGE 4

In this stage, participant A and participant B answer to some questions.

If you are a participant C, you do not make any decisions in this step. *If you are a participant A*, remember that, in stage 1, the participant B and the participant C answered to the following question: "Out of 3 participants A randomly selected in today's session, how many of these participants A will send 25 ECU to the participant B?". They answered to this question in each situation: North, West, East and South.

You must guess the answers of the participant B and of the participant C in your group.

If you are a participant B, remember that, in stage 1, the participant A and the participant C answered to the following question: "Out of 3 participants B randomly selected in today's session, if the participant A has sent 25 ECU, how many of these B participants will transfer the 25 ECU to another participant?". They answered to this question in each situation: North, West, East and South.

You must guess the answers of the participant A and of the participant C in your group.

How do the answers to these questions affect your earnings?

At the end of the session, for each role, one of the questions you answered during this stage will be randomly selected by the computer program. If your answer to this question is correct, you earn €1.

Example: Suppose you are a participant A and the randomly selected question is: "According to the participant C of your group, in the situation West, among 3 participants A randomly selected in today's session, how many of these participants A will send 25 ECU to the participant B?". If, in stage 1, the participant C of your group answered that according to him/her, in the situation West, x participant(s) A among the 3 participants A randomly selected to send 25 EMU to the participant B, then, your answer is correct if you answered "x".

I you have any questions raise your hand or press the red button on the side of your desk. We will answer your questions in private.

——————————— End of the fourth set of instructions ———————————

SECOND PART

In this part, the conversion rate is as follows: 10 EMU = €0.1.

There are fifteen periods.

In each period, you must choose the ECU allocation you prefer among nine allocations of ECU which will be proposed to you. An ECU allocations defines how many ECU you receive and how many ECU another participant X, randomly selected, receives.

Your earnings will be determined by one of your choices and by a choice of another participant Y, randomly selected. At the end of the session, a period will be randomly selected by the computer program, and the allocations chosen in this period determine your earnings:

- The allocation you have chosen during this period will be applied for you and for another participant X, randomly selected.
- The allocation that another randomly selected participant Y has chosen during this period will be applied for you and for him.

Your earnings for the second part is therefore the sum of what you receive in these two allocations selected.

END OF THE SESSION

At the end of the session, you will be informed of the decisions selected at random (yours and those other participants, if they affect you earnings) and of your final earnings.

Then, you will have to complete a final questionnaire.

At the end of the session, please sit and be quiet until an experimentalist invites you to go to the payment room. Take your computer tag and your payment receipt with you. Leave the instructions on your desk.
I you have any questions raise your hand or press the red button on the side of your desk. We will answer your questions in private.

——————————— End of the fifth set of instructions ———————————

DOCUMENTS DE TRAVAIL GREDEG PARUS EN 2020
GREDEG Working Papers Released in 2020

2020-01	Samira Demaria & Sandra Rigot *Taking on Board the Long-term Horizon in Financial and Accounting Literature*
2020-02	Gérard Mondello, Elena Sinelnikova & Pavel Trunin *Macro and Micro Implications of the Introduction of Central Bank Digital Currencies: An Overview*
2020-03	Gérard Mondello & Nissaf Ben Ayed *Agency Theory and Bank Governance: A Study of the Effectiveness of CEO's Remuneration for Risk Taking*
2020-04	Nicolas Brisset *Capital et idéologie : une critique*
2020-05	Giuseppe Attanasi, Alessandro Bucciol, Simona Cicognani & Natalia Montinari *Public The Italian North-South Divide in Perceived Dishonesty: A Matter of Trust?*
2020-06	Giuseppe Attanasi, Kene Boun My, Andrea Guido & Mathieu Lefebvre *Controlling Monopoly Power in a Double-Auction Market Experiment*
2020-07	Vera Barinova, Sylvie Rochhia & Stepan Zemtsov *How to Attract Highly Skilled Migrants into The Russian Regions*
2020-08	Guilhem Lecouteux *Welfare Economics in Large Worlds: Welfare and Public Policies in an Uncertain Environment*
2020-09	Raphaël Chiappini, Samira Demaria, Benjamin Montmartin & Sophie Pommet *Can Direct Innovation Subsidies Relax SMEs' Credit Constraints?*
2020-10	Giuseppe Attanasi, Samuele Centorrino & Elena Manzoni *Zero-Intelligence vs. Human Agents: An Experimental Analysis of the Efficiency of Double Auctions and Over-the-Counter Markets of Varying Sizes*
2020-11	Jean-Luc Gaffard *Entrepreneuriat et créativité : du détournement à la création de valeur*
2020-12	Michaël Assous, Muriel Dal Pont Legrand & Sonia Manseri *Samuelson's Neoclassical Synthesis in the Context of Growth Economics, 1956-1967*
2020-13	Frédéric Marty *Is the Consumer Welfare Obsolete? A European Union Competition Law Perspective*
2020-14	Charles Ayoubi, Sandra Barbosu, Michele Pezzoni & Fabiana Visentin *What Matters in Funding: The Value of Research Coherence and Alignment in Evaluators' Decisions*
2020-15	Giuseppe Attanasi, Claire Rimbaud & Marie-Claire Villeval *Guilt Aversion in (New) Games: the Role of Vulnerability*