



Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Exogenous Variation of Endowments and Beliefs in a Dictator Game

Pierpaolo Battigalli, Giovanni Di Bartolomeo
and Stefano Papa

Working Paper n. 725

This Version: September 2026

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
<http://www.igier.unibocconi.it>

The opinions expressed in the working papers are those of the authors alone, and not those of the Institute, which takes non institutional policy position, nor those of CEPR, NBER or Università Bocconi.

Exogenous Variation of Endowments and Beliefs in a Dictator Game*

Pierpaolo Battigalli[†] Giovanni Di Bartolomeo[‡] Stefano Papa[§]

September 2026

Abstract

We study sharing in a dictator game with asymmetrically observed random initial allocations. The design exogenously varies both the opportunity cost of sharing and the probability of receiving one of two possible endowment realizations. The treatment successfully shifts recipients' first-order beliefs and dictators' second-order beliefs. We interpret behavior through two channels: a belief-independent equal-sharing norm and a belief-dependent guilt-aversion motive. Conditional on the high-endowment realization, sharing is higher in the high-expectations treatment, consistent with guilt aversion. Overall, the evidence suggests that prosocial behavior reflects both norm compliance and concerns about disappointing others.

Keywords: sharing norm, opportunity cost, first and second-order beliefs, guilt aversion.

JEL classification codes: A13; C91; D01; D64.

*We thank Carlo Andreatta, Federico Golonia, Sem Manna, Elena Manzoni, Silvia Meneghesso, and Alessandro Stringhi for useful comments. Our experimental part was designed in 2021 and conducted in 2022 at the University of Rome Tor Vergata. The hypotheses of the experiment were not pre-registered. However, our hypotheses were articulated before the experiment.

[†]Department of Decision Sciences, IGIER, and BIDSa, Bocconi University, Italy; pierpaolo.battigalli@unibocconi.it.

[‡]Department of Economics and Law and CIMEO, Sapienza University, Italy; IEL, Corvinus University, Hungary; Department of Economics, University of Antwerp, Belgium; giovanni.dibartolomeo@uniroma1.it.

[§]Department of Economics and Finance, University of Rome Tor Vergata, Italy; CIMEO, Sapienza University of Rome, Italy; stefano.papa@uniroma2.it

1 Introduction

Dictator games have long been a workhorse for studying prosocial behavior. Two broad motivations dominate the literature. The first is belief-independent and encompasses norm-based and distributional preferences—such as altruism and fairness—according to which individuals share because they value others’ pay-offs or because they wish to comply with an equal-split norm, regardless of what the recipient expects (Fehr and Schmidt, 1999, 2010; Gächter *et al.*, 2017).¹ The second is belief-dependent: under guilt aversion, dictators experience a psychological cost when they fail to meet what they believe the recipient expects, so sharing may depend on second-order beliefs about the recipient’s expectations (Battigalli and Dufwenberg, 2007).²

Empirically testing these motivations is difficult because beliefs and behavior often co-move *ex post*. A dictator who intends to share may also report high beliefs about what the recipient expects, while a dictator who intends to keep may report low beliefs. Such correlations are informative, but they do not by themselves establish whether beliefs causally affect sharing or merely rationalize planned behavior.³ The distinction is important because norm-based motives and belief-based motives have different implications across social contexts and, therefore, for policy design.

We address this challenge by combining a simple theoretical framework with a controlled experimental design that generates two sources of variation. The first is exogenous variation in the realized opportunity cost of sharing. The second is exogenous variation in dictators’ second-order beliefs, induced by changing the commonly known probability that the dictator faces a low- rather than a high-endowment allocation. The realized endowment is observed only by the dictator and is realized after belief elicitation. Recipients therefore form expectations relying on the commonly known probability of two feasible endowments, and dictators

¹For a detailed taxonomy and definition of social norms, see Krupka and Weber (2013). The concept of social norms may also overlap with belief-dependent motivations (e.g., d’Adda *et al.*, 2020).

²Battigalli and Dufwenberg (2007) proposed a portable model of belief-based guilt aversion, later extended to allow for incomplete information and non-equilibrium analysis (e.g., Attanasi *et al.*, 2016, 2025; Battigalli *et al.*, 2019).

³Beliefs and actions are often *ex post* correlated, regardless of which motivation drives generosity (Khalmetzki, 2016). To test these motivations, methods introducing exogenous variation in expectations are necessary (Vanberg, 2008; Khalmetzki, 2016; Di Bartolomeo *et al.*, 2019).

can form second-order beliefs about those expectations.

The theory highlights two channels. A social-norm channel links sharing to its opportunity cost: holding beliefs constant, sharing should be less frequent when it is more costly. A guilt-aversion channel predicts that dictators' choices respond to second-order beliefs: sharing should be more likely when dictators believe that recipients expect them to share, especially when keeping would leave the recipient with very little and therefore generate larger expected disappointment. We expect the manipulation to shift beliefs if recipients assign a sufficiently high probability to dictators following an equal-splitting norm and dictators form second-order beliefs that are qualitatively consistent with this expectation. A higher probability of the low-cost realization should then raise recipients' first-order beliefs and dictators' second-order beliefs.

Guided by these predictions, we implement a modified dictator game with random and asymmetrically observed endowments. The dictator's endowment is realized after belief elicitation, and the recipient does not observe the realized endowment. We manipulate the publicly known probability of facing a low- versus high-endowment game, which we expect to shift first- and second-order beliefs about sharing, while the realized endowment determines the material cost of sharing. We run the experiment with 512 undergraduate participants at Tor Vergata University of Rome.

Our main contribution is methodological and substantive. Methodologically, we show how a change in the commonly known distribution of hidden endowments can be used to shift expectations while still allowing behavioral comparisons conditional on the realized endowment. Substantively, we provide evidence consistent with a joint role for belief-independent and belief-dependent motives. We are careful, however, not to claim that the treatment isolates a pure direct manipulation of beliefs. The treatment consists of a commonly known variation in the probability distribution of endowments, which turns out to exogenously shift second-order beliefs, although individual beliefs may still partly reflect projection or intended behavior. For this reason, our interpretation relies on the combination of three pieces of evidence: the manipulation check on beliefs, the reduced-form treatment comparison conditional on the realized endowment, and the individual-level association between second-order beliefs and sharing.

Our work relates to evidence that initial endowments, relative payoffs, and the cost of prosocial behavior shape giving. Korenok *et al.* (2012) show that dictators’ transfers are sensitive to the recipient’s endowment, consistent with inequality aversion, while Korenok *et al.* (2014) show that refraining from taking is not behaviorally equivalent to giving. Nguyen (2019) further documents that dictators give less when they learn that other dictators receive substantially larger endowments, consistent with self-serving fairness considerations. More broadly, Cox’s (2004) triadic design, and its extensions by Di Bartolomeo and Papa (2016), highlight the importance of distinguishing conditional motives, such as trust and reciprocity, from unconditional other-regarding preferences. We contribute to this literature by showing that the commonly known distribution of hidden endowments can shift expectations, thereby providing a way to separate the role of realized opportunity costs from belief-dependent motives in dictator giving.

We are also closely connected to the literature on guilt aversion in dictator games. Ellingsen *et al.* (2010) and Khlametski *et al.* (2015) emphasize both the conceptual difficulty of distinguishing belief-dependent guilt aversion from other social-preference models and the practical challenges involved in eliciting beliefs. Hauge (2016) is particularly relevant because it separates two channels through which recipients may indirectly affect dictator giving: recipients’ expectations about what dictators will give and recipients’ moral standards about what dictators ought to give. He shows that dictators give more when they condition their choices on higher recipient expectations, but also when recipients state higher moral standards. This distinction is important for our interpretation, since changes in perceived normative standards could mimic, or operate alongside, guilt aversion. Relative to Hauge (2016), our design does not directly condition choices on recipients’ stated expectations or moral standards. Instead, we shift beliefs through the commonly known distribution of hidden endowments and compare behavior conditional on the dictator’s realized material incentives. Stringhi (2026) compares guilt aversion and inequity aversion in a trust minigame, showing that the two motives can generate opposite predictions when the co-player’s material payoffs are varied. Although his setting differs from ours, the paper is close in spirit because it uses a theory-driven payoff manipulation to separate belief-dependent guilt from outcome-based fairness concerns.

Methodologically, we build on Di Bartolomeo *et al.* (2019), who exploit a partner-switching mechanism with asymmetric information and experimentally varied re-matching probabilities to induce exogenous variation in expectations.⁴ Relative to this literature, our contribution is to study the same broad identification problem in a dictator-game environment with hidden endowments. Rather than varying the recipient’s payoff directly or using communication, we generate variation in expectations through the commonly known distribution of hidden endowments. This allows us to test for a belief-dependent guilt-aversion channel while conditioning on the dictator’s realized material incentives.

The interaction between guilt aversion and social norms in dictator games is studied by Danilov *et al.* (2021), who use a different experimental design motivated by a different theoretical model. While our framework contrasts guilt aversion with an injunctive equal-splitting norm, their model distinguishes guilt aversion from conformity to a descriptive norm, namely the desire to choose a transfer close to the behavior of others in the same situation. They disentangle these motives by providing dictators with exogenous information about both the recipient’s expectation and the average transfer of other dictators. Our contribution is therefore not to show that norms and guilt aversion can coexist *per se*, but to show that variation in the commonly known distribution of hidden endowments shifts expectations and thereby provides a test of guilt aversion conditional on realized material incentives.

The rest of the paper is organized as follows. Section 2 introduces the theoretical framework and the hypotheses guiding the experimental design. Section 3 describes the experimental implementation and the sources of variation used to separate belief-dependent from norm-driven motives. Section 4 presents the analysis and main findings. Section 5 provides some robustness and specification checks. Section 6 discusses identification and alternative interpretations. Section 7 reports pooled tests of exploratory gender heterogeneity. Section 8 concludes. The Appendix contains details about belief elicitation, sensitivity analyses, gender-specific estimates, and conditions on the primitives of our model that yield the auxiliary assumption of Section 2.

⁴This methodology was also used in Ciccarone *et al.* (2020) and Di Bartolomeo *et al.* (2024). See Sections 2 and 3 for the details of our experimental design.

2 Experimental design and hypotheses

This section describes the experimental design (2.1), develops a simple theoretical model with norm-driven and guilt-driven types (2.2), and presents the experimental hypotheses (2.3).

2.1 A dictator game with hidden initial allocations

We analyze a *Share-or-Keep Dictator Game* with random initial endowments (Battigalli *et al.*, 2025). Payoffs represent monetary payments rather than utilities, acknowledging that individual decisions might be influenced by other-regarding preferences, such as social norms and feelings of guilt. In this setup, given the observed realization $x \in \{x_L, x_H\}$ (with $x_L < x_H$) of a random initial allocation $\mathbf{x}$ for the dictator and $M - \mathbf{x}$ for the recipient, the dictator faces a binary decision: to *Keep* the initial allocation, preserving the status quo, or to *Share*, redistributing the resources equally between the two parties.

Specifically, Chance splits M tokens as endowments of the dictator and the recipient, assigning x_L or x_H tokens to the dictator with, respectively, probability p or $1 - p$, and the rest to the recipient. Next, the dictator can either *Share* the total number of tokens (M) equally or *Keep* the initial allocation (status quo). The distribution of the random endowment $\mathbf{x}$ is commonly known, but its realization x is observed *only* by the dictator (asymmetric information), and this is common knowledge. The dotted line emphasizes that if the dictator shares, the recipient cannot infer the dictator’s initial endowment from observing their own payoffs.⁵

Of course, the own-payoff-maximizing decision of the dictator is to always keep, regardless of the observed value x . However, incorporating other-regarding preferences shifts the focus to how social norms and anticipated guilt from keeping might shape the dictator’s decision. The material opportunity cost of sharing, $x - M/2$, increases linearly in x . Therefore, with a slight abuse of language, we often refer to x as the (opportunity) cost of sharing. Furthermore, for the sake of brevity,

⁵This reduces the potential impact of self-image concerns (Benabou and Tirole, 2006; Murnighan *et al.*, 2001).

we refer to the “subgame” where the dictator initially holds x_H as the x_H -Game, while the other is called the x_L -Game.⁶

Our **experimental design** involves two treatments, which differ in the value of the probability $p = \Pr(\mathbf{x} = x_L)$ of playing the low-endowment (*low-cost*) game. In **treatment 1** (t_1), we set $p = 0.75$. In **treatment 2** (t_2), we set $p = 0.25$. The other parameters are: $M = 40$, $x_L = 24 > M/2$, and $x_H = 40 = M$.

To anticipate, our main experimental hypotheses are that (a) sharing is more frequent and recipients’ first-order beliefs are higher in t_1 than in t_2 , and (b) dictators’ second-order beliefs reflect this.

2.2 Social norms and guilt aversion: A simple model

We assume that a fraction $q_n \in (0, 1)$ of agents in the dictator role, referred to as “norm-driven (n -)dictators,” dislike violating an equal-sharing norm. The remaining fraction $q_g = 1 - q_n$ of agents are “guilt-driven (g -)dictators,” who are instead motivated by guilt aversion. Furthermore, we assume that within each subgroup, agents are heterogeneous, i.e., they differ in the degree to which they care about adhering to social norms or about guilt.⁷

The parameterized utility function of an n -dictator is

$$u_n(m_D, m_R) = m_D - \theta_n \mathbf{1}_{\{m_D > m_R\}}, \quad (1)$$

where m_i is the realized monetary payoff of agent i in $\{D, R\}$,⁸ $\mathbf{1}_{\{m_D > m_R\}}$ is the indicator function of (dictator-biased) unequal distribution, and parameter $\theta_n \geq 0$ gauges the psychological cost of violating the social norm of equal sharing.

⁶Technically, these are not true subgames because the recipient’s terminal nodes following the two possible endowment realizations belong to the same information set. Thus, the recipient does not observe the realized endowment even ex post, which can matter for belief-dependent utility.

⁷The assumption of two heterogeneous groups is introduced for expositional simplicity. The more complex alternative would be to consider bidimensional types.

⁸The label D refers to the dictator, the agent who decides how to allocate the payoff, while R refers to the recipient, the passive player who receives the allocation. These labels are used as subscripts to distinguish the payoffs of the two players.

The parameterized utility function of a g -dictator is

$$u_g(m_D, m_R, \alpha_R) = m_D - \theta_g [E_\alpha(\mathbf{m}_R) - m_R]^+, \quad (2)$$

where $[y]^+ = \max\{0, y\}$, α represents the recipient's first-order probabilistic belief about the variables that determine their payoff, and parameter $\theta_g \geq 0$ gauges the dictator's guilt sensitivity, i.e., the individual aversion to letting down (disappointing) the recipient. Note that – as is now standard in psychological game theory – the dictator's utility depends on the recipient's belief, which is unknown to the dictator, who has to consult their second-order belief β to determine the subjectively optimal choice.⁹

For n -dictators, the indifference condition $x - \theta_n = M/2$ yields threshold

$$\hat{\theta}_n(x) = x - \frac{M}{2}, \quad (3)$$

which is the parameter value above (below) which a dictator chooses to share (not share). Note that $\hat{\theta}_n(x_H) - \hat{\theta}_n(x_L) = x_H - x_L > 0$. As long as the interval $(x_L, x_H]$ contains a positive mass of n -types, increasing the endowment from x_L to x_H strictly decreases the sharing rate (which moves in the opposite direction of the threshold) in the subpopulation of n -dictators. This holds independently of agents' beliefs and chance probability p .

In our design, $\hat{\theta}_n(x_H) = 40 - 20 > \hat{\theta}_n(x_L) = 24 - 20$; thus, *for norm-driven dictators the sharing rate is higher when the material opportunity cost of sharing, parameterized by x , is lower*. We will refer to this as the **social-norm channel**.

Next we analyze the subpopulation of g -dictators. We refer to the effects of exogenous variations on the behavior of g -dictators as the **guilt-aversion channel**. The effect of guilt aversion is isolated by looking at how second-order beliefs β affect behavior for each initial endowment x . However, we also have to analyze the effect of a change in endowment. Since it is commonly known that the realized endowment x is observed only by the dictator, recipients' first-order beliefs (FOBs) cannot depend on it, and it makes sense to assume that the dictators' second-order

⁹See Battigalli and Dufwenberg (2022) for a comprehensive review and discussion of psychological game theory.

beliefs (SOBs) are independent of x as well. Accordingly, our analysis of how dictators' choices vary with x holds *second-order beliefs* β constant. We are going to show that, unlike n -dictators, *for g -dictators the sharing rate is increasing in x* . Here we prove this under a simplifying assumption, but the result is general (see Battigalli *et al.* 2025).

Let $\bar{m}(\beta) := E_\beta(E_\alpha(\mathbf{m}_R))$ denote the dictator's expectation of how much the recipient expects to get according to SOB β . With this, under the simplifying assumption that *the dictator is certain that* $E_\alpha(\mathbf{m}_R) > M - x$ (i.e., that keeping induces positive disappointment), the expected utility of keeping is $x - \theta_g[\bar{m}(\beta) - (M - x)]$, with $\bar{m}(\beta) > (M - x)$. Since sharing cannot disappoint, its utility is simply $M/2$. Thus, the indifference condition $\theta_g[\bar{m}(\beta) - (M - x)] = x - M/2$ yields threshold

$$\hat{\theta}_g(\beta, x) = \frac{x - \frac{M}{2}}{\bar{m}(\beta) - (M - x)} = \frac{x - \frac{M}{2}}{x - (M - \bar{m}(\beta))}. \quad (4)$$

The equation shows that if the dictator's second-order expectation $\bar{m}(\beta)$ is low, i.e., close to $(M - x)$, the disappointment that the dictator expects to inflict is low and guilt sensitivity must be high to induce the dictator to share. Thus, *the sharing rate* (which moves in the opposite direction of the threshold) *is increasing in $\bar{m}(\beta)$* . Furthermore, $\lim_{\bar{m}(\beta) \searrow (M - x)} \hat{\theta}_g(\beta, x) = +\infty$.

In (4), the numerator represents the material (opportunity) cost of sharing, while the denominator is proportional to the expected psychological cost of keeping (i.e., of disappointing the recipient). An increase in x raises both costs. Thus, *these two effects partially cancel out*. However, unlike $\hat{\theta}_n(x)$, threshold $\hat{\theta}_g(\beta, x)$ is *decreasing* in x . To see this, note that $\bar{m}(\beta) \leq M/2$, because the recipients' most optimistic expectation is that the dictator shares with probability 1 independently of x , i.e., that they certainly get $M/2$. Thus, $M - \bar{m}(\beta) \geq M/2$, making the numerator in the right-hand side of (4) weakly larger than the denominator. This implies that an increase in x increases the numerator proportionally less than the denominator, and the fraction decreases.

Battigalli *et al.* (2025) show that the same conclusion also holds if the dictator is *not* certain to disappoint the recipient, and under more general assumptions about the psychological utility function of guilt-driven agents.

Observation 1 (Belief boundaries) *The lowest value of second-order expectation $\bar{m}(\beta)$ is attained when, according to SOB β , recipients are certain that dictators never share, which yields the recipient's random payoff $M - \mathbf{x}$. The highest value of $\bar{m}(\beta)$ is attained when, according to β , recipients are certain that dictators always share, which yields payoff $M/2$ with certainty. In our design, this implies treatment-dependent bounds on $\bar{m}(\beta)$. Letting $\bar{m}(\beta|t)$ denote the value of $\bar{m}(\beta)$ in treatment $t \in \{t_1, t_2\}$, we have $\bar{m}(\beta|t_1) \in [12, 20]$ and $\bar{m}(\beta|t_2) \in [4, 20]$. Note that the lower bound is not zero and differs across treatments because the probability of playing the 24-Game, in which $M - x = 16$, is not zero.*

As explained above, the social-norm channel and the guilt channel make sharing behavior react in opposite ways to an increase in dictators' endowment, with the former implying less sharing due to the higher opportunity cost. However, the guilt channel is the net result of two effects that partially cancel out. Therefore, it is plausible to assume that, if the fraction of dictators affected by the social-norm motivation is not much smaller than the complementary fraction of dictators affected by guilt aversion, the aggregate effect of an increase in dictators' endowment is to decrease sharing:

Auxiliary Hypothesis. Under plausible assumptions about the distribution of types, we expect that on average and keeping beliefs fixed, *the social-norm channel prevails, and the sharing rate is lower when the dictators' endowment is high*. Furthermore, we assume that recipients believe that this is the case and therefore have *higher FOBs in treatment t_1* , with low opportunity cost of sharing. Finally, we assume that dictators believe that FOBs are higher in treatment t_1 , so that their *SOBs are also higher in treatment t_1* .

In the Appendix we provide conditions on the primitives of the model implying the prevalence of the social-norm channel. The substantive further assumption is that beliefs qualitatively agree with such prevalence. (However, we are not imposing the equilibrium assumption that beliefs are correct.)

Observation 2 *In each treatment of our design, beliefs $\bar{m}(\beta)$ are fixed, but the threshold for guilt aversion (4) varies depending on the endowment observed by the*

dictator. That is

$$\begin{aligned}\hat{\theta}_g(\beta, x_H) &= \frac{40 - 20}{[40 - (40 - \bar{m}(\beta))]^+} = \frac{20}{\bar{m}(\beta)} \\ \hat{\theta}_g(\beta, x_L) &= \frac{24 - 20}{[24 - (40 - \bar{m}(\beta))]^+} = \frac{4}{[\bar{m}(\beta) - 16]^+},\end{aligned}$$

with $\hat{\theta}_g(\beta, x_L) = +\infty$ if $[\bar{m}(\beta) - 16]^+ = 0$, i.e., if $\bar{m}(\beta) \leq 16$. The effects of guilt aversion are always triggered when the dictator is in the high-endowment condition (x_H), whereas in the low-endowment condition (x_L), they will only be triggered if $\bar{m}(\beta) > 16$.

From the above observation, we can state the following. In our design, beliefs $\bar{m}(\beta)$ are independent of the endowment within each treatment, but they vary across treatments due to differences in the distribution of the random endowment $\mathbf{x}$. If the social-norm channel is dominant, and this is internalized in beliefs, we should expect $\bar{m}(\beta|t_1) > \bar{m}(\beta|t_2)$, and guilt aversion will imply corresponding differences in behavior across treatments. In the 40-Game:

$$\hat{\theta}_g(\beta, x_H|t_1) = \frac{20}{\bar{m}(\beta|t_1)} < \frac{20}{\bar{m}(\beta|t_2)} = \hat{\theta}_g(\beta, x_H|t_2).$$

Similarly, in the 24-Game:

$$\hat{\theta}_g(\beta, x_L|t_1) = \frac{4}{[\bar{m}(\beta|t_1) - 16]^+} \leq \frac{4}{[\bar{m}(\beta|t_2) - 16]^+} = \hat{\theta}_g(\beta, x_L|t_2),$$

where the equality holds if $\bar{m}(\beta|t_1) \leq 16$. In particular, when the dictator plays the 24-Game, guilt aversion is not triggered by our exogenous variation in beliefs, unless dictators' SOBs are very high in the high-expectations treatment t_1 .

The inequalities above identify the guilt-aversion channel. Our analysis suggests that this channel should more strongly affect behavior in the high-endowment condition (40-Game), than in the low-endowment condition (24-Game), and in the latter it may have a negligible effect.

2.3 Experimental hypotheses and identifying variation

Based on the theoretical model, we derive several testable predictions. When the dictator's endowment (and thus the opportunity cost of sharing) increases, the two motivational channels move in opposite directions. The social-norm channel predicts a lower sharing rate because complying with the equal-sharing norm becomes more costly. The guilt-aversion channel predicts a weakly higher sharing rate because keeping is expected to inflict more disappointment on the recipient. Under plausible assumptions on the distribution of types, we expect the social-norm channel to dominate on average, so that higher sharing costs reduce average sharing rates. Moreover, a higher probability of facing the low-cost allocation should raise both first-order beliefs (FOBs) and second-order beliefs (SOBs). If guilt aversion is present, this probability should also affect sharing through the belief-dependent channel.

The first step is therefore a manipulation check. Recipients should believe that dictators are more likely to share when the low-cost/endowment condition is more likely. Since the information structure is common knowledge, dictators should internalize recipients' reasoning when forming their own second-order beliefs. We test:

$$H1a : \text{FOB}(t_1) > \text{FOB}(t_2) \tag{5}$$

$$H1b : \text{SOB}(t_1) > \text{SOB}(t_2), \tag{6}$$

where $\text{FOB}(t)$ and $\text{SOB}(t)$ respectively denote the probability with which recipients expect dictators to share and the dictators' expectation of this probability in treatment $t \in \{t_1, t_2\}$.

The verification of (5) and (6) supports the claim that the experimental manipulation changes beliefs in the intended direction. It is also a necessary ingredient for our behavioral test of the belief-dependent channel. Without such a belief shift, any treatment difference in sharing would be more naturally interpreted as a belief-independent treatment effect.

Our primary behavioral test of the guilt-aversion channel is the comparison across treatments conditional on the high-endowment realization. In the 40-Game,

choosing *Keep* gives the recipient zero. Hence, if recipients assign positive probability to sharing, keeping yields high disappointment. Assuming that (6) holds, guilt aversion predicts:

$$H2a : \text{share}(t_1|x_H) > \text{share}(t_2|x_H). \quad (7)$$

This comparison holds the realized endowment fixed and varies beliefs. It is therefore the cleanest reduced-form test of whether higher second-order expectations are associated with more sharing when expected disappointment is large.

We also examine the corresponding comparison in the 24-Game:

$$H2b : \text{share}(t_1|x_L) \geq \text{share}(t_2|x_L). \quad (8)$$

This prediction is informative but weaker. In the low-endowment condition, keeping gives the recipient 16 tokens. Therefore, guilt aversion is activated only if dictators believe that recipients expect more than 16. Equivalently, in the notation of the model, the relevant threshold is $\hat{\theta}_g(\beta, x_L) = 4/[\bar{m}(\beta) - 16]^+$, which is infinite when $\bar{m}(\beta) \leq 16$. Thus, failure to find a strong treatment difference in the 24-Game would not refute the guilt-aversion interpretation.

Hypotheses (9) and (10) follow from the social-norm channel. The estimation of this effect relies on comparing sharing rates under the low- and high-endowment conditions within each treatment. Formally:

$$H3a : \text{share}(t_1|x_L) > \text{share}(t_1|x_H), \quad (9)$$

$$H3b : \text{share}(t_2|x_L) > \text{share}(t_2|x_H). \quad (10)$$

These comparisons hold the treatment-specific beliefs fixed while varying the realized opportunity cost of sharing. They should therefore pick up the net effect of the endowment change. Under the auxiliary hypothesis, the social-norm channel dominates the countervailing guilt-aversion channel, so the sharing rate should be higher when the dictator's endowment is low.

Finally, the model suggests that the within-treatment effect of reducing the opportunity cost should be smaller in the high-expectations treatment, because guilt aversion offsets the cost effect more strongly when second-order expectations

are higher. This yields the difference-in-differences prediction:

$$H4 : [\text{share}(t_2|x_L) - \text{share}(t_2|x_H)] > [\text{share}(t_1|x_L) - \text{share}(t_1|x_H)], \quad (11)$$

which is equivalent to hypothesizing a higher treatment effect in the high-cost condition.

We first test whether the treatment shifts beliefs, using FOBs and SOBs as manipulation checks. We then study sharing decisions conditional on the realized endowment. The comparison across treatments within the x_H -Game is our main reduced-form test of the belief-dependent channel, because material incentives are held fixed while beliefs vary. The comparison across endowments within treatment is instead used to assess the cost-sensitive, belief-independent channel. Finally, individual-level regressions using SOBs are interpreted as mechanism evidence: they show whether dictators who believe recipients expect more sharing are more likely to share, but they are not treated as sufficient for causal identification.

3 Experimental procedure

The experiment was conducted at Tor Vergata University (Rome) in May 2022 and involved 512 undergraduate students across 16 sessions, recruited through an online recruitment system. The entire experiment was computerized and implemented using the *z-Tree* software (Fischbacher, 2007). Upon arrival at the laboratory, participants were randomly assigned to individual and isolated computer terminals. Two research assistants distributed the instructions and ensured that participants followed the procedures correctly. Before beginning the main tasks, all subjects completed a brief comprehension questionnaire to verify their understanding of the instructions.

Each session consisted of 8 rounds (with perfect stranger matching) in a within-subject design. Our design is built on the simple game described in Figure 1. We always let $M = 40$, $x_L = 24$, and $x_H = 40$. A high opportunity cost of sharing characterizes the 40-Game, i.e., the equal payoff distribution costs 20 tokens to the dictator in the 40-Game and only 4 tokens in the 24-Game.

Each round implemented the following five-stage procedure.

1. Role assignment. Player roles A and B are randomly assigned, and pairs are formed.
2. Probability of distribution (treatment). Participants are informed about the probability p that the round will be a 24-Game (i.e., that the dictator’s endowment will be $x_L = 24$). In each round, p can be either 0.75 (t_1) or 0.25 (t_2). The same value of p applies to all pairs in a session-round. Each session includes four rounds under each treatment. In sequence A, t_1 is implemented in rounds 1, 3, 5, and 8; in sequence B, t_1 is implemented in rounds 2, 4, 6, and 7. The two complementary sequences are randomly assigned across sessions, with eight sessions using each sequence.
3. Belief elicitation. Recipients’ FOBs and dictators’ SOBs are elicited before the endowment realization. This stage has two parts: (a) first-order beliefs: each recipient predicts the dictator’s action; (b) second-order beliefs: each dictator predicts the recipient’s first-order belief. Recipients choose among five ordered probability categories and are paid according to a quadratic scoring rule. Dictators receive 0.50 tokens for correctly identifying the category selected by their counterpart. Appendix A1 reports the response categories and payment schedule.
4. Endowment assignment. Endowments are randomly assigned according to p . Only dictators observe the realized endowment (asymmetric information).
5. Dictator’s action. Dictators choose whether to share equally (20–20) or to keep the randomly determined status quo.

At the end of each session, one round was randomly selected for payment based on the allocation decision. Belief elicitation was instead incentivized in all remaining rounds.¹⁰ All the game payoffs were described in terms of “tokens,”

¹⁰The belief-elicitation procedure is adapted from Vanberg (2008) and proceeds as follows (see the Appendix for details). First, each recipient states a prediction about the dictator’s action. Second, each dictator reports her prediction of the recipient’s first-order belief. Beliefs were incentivized in rounds that were not used to determine monetary payoffs. Similar elicitation schemes are widely used in the experimental literature (e.g., Di Bartolomeo *et al.*, 2019; Ciccarone *et al.*, 2020; Di Bartolomeo *et al.*, 2023a, 2023b, 2024; Charness *et al.*, 2024). This procedure reduces incentives to hedge between actions and belief reports. However, as discussed by Cox *et*

with 1 token = 0.50 euros. In addition, each subject received a fixed show-up fee of 3 tokens.

Two features of the elicitation procedure are important for interpretation. First, beliefs are elicited before the realized endowment is observed by the dictator. This timing helps separate the treatment-induced beliefs from the realized opportunity cost of sharing. Second, the elicited SOB is a coarse proxy for the dictator’s expectation of the recipient’s first-order belief: dictators guess which of five categories the recipient selected. We therefore use individual SOBs as a mechanism variable rather than as a perfect cardinal measure of $\bar{m}(\beta)$.

4 Experimental results

Our results are displayed in Table 2, which reports the elicited average FOBs and SOBs (about how likely the dictator is to share) and sharing rates. Experimental outcomes are structured by treatment (t_1 and t_2 ; rows) and endowment (24-vs. 40-Game; columns). The table also reports the standard deviations below the sharing rates. In column (a), each treatment cell contains 1,024 recipient-round observations; in column (b), it contains 1,024 dictator-round observations. Each treatment-by-endowment cell in columns (c)–(f) contains 512 dictator-round observations. These counts refer to repeated role-round observations, not to distinct participants. Overall, average sharing rates vary between 35% and 46% across treatments and endowments. Average FOBs and SOBs range between 0.35 and 0.49.

al. (2015), hedging concerns cannot be entirely ruled out when multiple incentivized tasks are combined.

Table 2: FOBs, SOBs, and sharing rates

	FOBs	SOBs			Sharing rate	
	All	All	40-G.	24-G.	40-G.	24-G.
	(a)	(b)	(c)	(d)	(e)	(f)
(1) High prob. to	48%	48%	49%	48%	42%	46%
play 24-Game (t_1)	0.31	0.31	0.31	0.31	0.49	0.50
(2) Low prob. to	35%	36%	36%	36%	35%	41%
play 24-Game (t_2)	0.31	0.31	0.31	0.31	0.48	0.49

The manipulation of the probability of playing the low-endowment game (24-Game) works as intended (*H1*): when the 24-Game is more likely to be played, both FOBs and SOBs that the dictator will share are substantially higher than when it is less likely. On average, subjects are less confident that dictators will share when the material opportunity cost of sharing is more likely to be high. As explained, this result is consistent with the existence of a (predominant) social-norm channel.

In columns (a) and (b) of Table 2, we respectively observe 48% vs. 35% for FOBs and 48% vs. 36% for SOBs.¹¹ The percentages in Table 2 are descriptive. Formal inference accounts for assignment at the session-round level and dependence within sessions, as detailed in Tables 4 and 5.

The exogenous variation in probability p of the low endowment allows us to compare, *ceteris paribus*, two different levels of SOBs and thereby test for the guilt aversion channel. As implied by the theoretical model, we expect this effect to be larger when keeping yields a payoff of zero to the recipient (*H2a*). By contrast, in our model the effects of guilt aversion are small or negligible when keeping grants 16 tokens to the recipient, because in this case it is unlikely that such a choice will generate (and be expected to generate) substantial disappointment in the recipient (*H2b*).

Our empirical strategy builds on the exogenous variation in beliefs induced by our experimental design: we isolate the guilt-aversion channel and compare

¹¹We also compare SOBs between dictators who later play the 40- and 24-Game (columns (c) and (d)). As expected, these are unaffected by the randomly selected subgame, since expectations are elicited before dictators learn which subgame they will play (tests omitted).

sharing rates within columns (e) and (f), which correspond to hypotheses *H2a* and *H2b*, respectively. Sharing is higher in the high-expectations treatment under both endowments. When keeping leaves the recipient with zero (*H2a*), the raw difference is 7 percentage points (42% vs. 35%). In the low-endowment condition (*H2b*), the raw difference is 5 percentage points (46% vs. 41%). These percentages do not by themselves provide valid treatment-level inference because the probability treatment is common to all participants in a session-round. The formal session-round-cell estimates and exact randomization tests are reported in Table 5.

Summing up, *H2a* is our sole primary behavioral test and provides the primary evidence in favor of the belief-dependent channel: in the condition where keeping is most likely to disappoint the recipient, sharing is higher in the high-expectations treatment. The two-sided exact randomization test is marginal ($p = 0.071$), so we interpret this pattern as suggestive evidence consistent with guilt aversion. The *H2b* point estimate is smaller ($p = 0.099$ in the two-sided exact test), but the difference between the two treatment effects is not statistically significant. The weaker theoretical status of *H2b* follows from the fact that keeping leaves the recipient with 16 tokens, which is less likely to generate disappointment.

Next, we turn to non-belief-dependent motivations, i.e., the social-norm channel (*H3*). Our exogenous variation in the endowment (hence, in the opportunity cost of sharing) allows us to observe its effect by comparing sharing rates within rows, i.e., holding beliefs fixed within each treatment. As argued above, dictators facing a lower material cost of sharing should be more likely to choose *Share*—both when SOBs are high and when they are low—because complying with the equal-sharing norm is less demanding. However, changes in the opportunity cost also activate the (non-predominant) guilt-aversion channel, which may partially offset the social-norm one. Since the guilt-aversion channel is stronger when expectations are high, we expect the net impact of a fall in the opportunity cost of sharing to be smaller in the high-expectations treatment t_1 than in the low-expectations treatment t_2 (diff-in-diff hypothesis *H4*).

Table 2 shows that, in both treatments, the sharing rate is higher when the initial endowment is lower: it rises from 42% to 46% in t_1 and from 35% to 41% in t_2 . Taken together with the results for *H1*, these directional patterns support the

presence of a (predominant) social-norm channel. Moreover, the effect is larger (6 p.p. vs. 4 p.p.) when the probability of being in the 24-Game group is smaller, as predicted by the theory: in this treatment (t_2), the mitigating effect of the countervailing guilt-aversion channel is lower. However, the formal fixed-effects estimates reported in Table 6 are imprecise. The estimated effect of the 40-Game relative to the 24-Game is -0.78 percentage points in the high-expectations treatment ($p = 0.851$) and -5.86 percentage points in the low-expectations treatment ($p = 0.319$). In the separate primary session-round-cell specification reported in Table 5, the difference-in-differences estimate is 4.56 percentage points ($p = 0.386$), with a two-sided exact randomization p -value of 0.399 . Thus, both approaches yield results that are directionally consistent with $H3$ and $H4$, but the data do not establish either the within-treatment endowment effects or a difference between them.

Our experimental design generates exogenous variation in second-order beliefs (SOBs), effectively creating a high- versus low-expectations environment. Beyond this treatment-level shift, we also elicit each dictator’s individual SOB. Exploiting this richer, individual-level measure allows us to move beyond session averages and examine sharing behavior at a finer level of detail. The corresponding results are reported in Table 3.

Table 3 reports estimates from an individual fixed-effects linear probability model in which the dependent variable is the dictator’s action—equal to 100 if *Share* was chosen, and 0 otherwise. We use a linear probability model rather than probit or logit because, in our setting, sharing rates fall in a relatively narrow mid-range interval (35–46%), so its coefficients can be read directly as percentage-point changes in the probability of choosing *Share*.¹² The key explanatory variables are the dictator’s SOB, the endowment condition (24 or 40), and their interaction, with an additional control for the round in which the decision was made. The specification allows for a direct interpretation of the coefficients as marginal effects, i.e., p.p. changes in the probability of choosing *Share*. SOBs and the round variable have been standardized, and the 24-Game is the reference endowment condition.

¹²Coefficients can be interpreted as marginal effects—percentage point changes in the probability of choosing *Share*. SOBs and the round variable are standardized; the reference category corresponds to average belief (42.3%) and average round (4.5).

Table 3: Sharing behavior regression

	coef.	s.e. rob.	t	p-value	95% CI	
	(a)	(b)	(c)	(d)	(e)	(f)
(1) SOBs	5.69	1.85	3.08	0.002	2.06	9.33
(2) endowment	-5.68	2.21	-2.57	0.010	-10.01	-1.35
(3) (1) $\times$ (2)	1.77	2.26	0.78	0.433	-2.67	6.21
(4) round	-1.55	0.93	-1.66	0.098	-3.38	0.28
(5) constant	44.00	1.10	39.87	< 0.001	41.83	46.17

Note: The specification includes individual fixed effects and a standardized round control. Standard errors are clustered at the individual level (2,048 dictator-round observations from 512 participants, each contributing four dictator decisions). SOBs are standardized over the dictator sample, and the endowment indicator equals one for the 40-Game. Coefficients are expressed in percentage points. Appendix Table A2 reports the same specification with standard errors clustered at the session level.

SOBs and the endowment map naturally into our two mechanisms. A higher SOB means that the dictator believes the recipient expects sharing with higher probability, so guilt aversion predicts a positive coefficient on SOBs. The endowment dummy equals 1 when the dictator’s endowment is high (40), i.e., when sharing is more costly (the opportunity cost of choosing *Share* is larger). Cost-based motives—including compliance with an equal-split norm when it is less demanding—therefore predict a negative coefficient on the endowment dummy.

A key caveat is that SOBs are not a direct measure of the dictator’s expectation of the recipient’s disappointment. They measure the dictator’s expectation of the probability assigned by the recipient to the sharing action, but whether the Keep action is disappointing depends on how much the recipient obtains, which in turn depends on the endowment observed only by the dictator: for fixed FOBs, the recipient is more disappointed when the endowment that the dictator keeps is high.

For this reason, the interaction term should be read as capturing whether the belief effect is amplified when the endowment is high. A positive interaction is consistent with the idea that, at higher endowments, keeping is more likely to fall short of what the recipient expects, which strengthens the guilt motive and can partially offset the larger opportunity cost of sharing.

Table 3 shows that higher SOBs are associated with a significantly greater probability of sharing. In the 24-Game, a one-standard-deviation increase in SOBs is associated with a 5.69 percentage-point increase in sharing (SE = 1.85, 95% CI [2.06, 9.33], $p = 0.002$). In the 40-Game, the corresponding effect is 7.46 percentage points (SE = 1.78, 95% CI [3.97, 10.96], $p < 0.001$). The difference between the two slopes is 1.77 percentage points (SE = 2.26, 95% CI [-2.67, 6.21], $p = 0.433$). Thus, SOBs are positively associated with sharing under both endowments, but the data do not establish that this association is stronger in the 40-Game.

The negative and significant coefficient on the high-endowment dummy indicates that a higher material opportunity cost of sharing reduces the propensity to share by 5.68 p.p. This pattern is consistent with the social-norm channel, although the coefficient captures the net effect of the endowment change rather than a direct measure of norm compliance. That is, holding SOBs and round at their mean values, the fitted values under the fixed-effects normalization are 44.0% for the low-cost endowment and 38.3% for the high-cost endowment.

Consistent with the presence of the guilt-aversion channel, the interaction term between SOBs and endowment is positive, although not statistically significant, suggesting that the influence of beliefs tends to become stronger when keeping gives less to the recipient.

Finally, there is a negative time trend, consistent with some learning or adaptation over rounds.

In sum, the data indicate that variation in the endowment (i.e., an increase in the opportunity cost) is associated with a decrease of about 5 p.p. in the probability of choosing to share. Moreover, individual SOBs are positively associated with sharing under both endowments, with estimated effects between 5.69 and 7.46 percentage points per standard deviation. This association is consistent with guilt aversion, but we do not treat it as stand-alone causal evidence because individual beliefs may partly reflect projection or intended behavior. The interaction

between SOBs and endowment is positive, as predicted, but it is not statistically significant.

5 Robustness and specification checks

The main treatment comparisons use session-round-endowment-cell means, session-clustered standard errors, and exact randomization inference based on the assignment of treatment sequences to sessions. The individual-level analyses use linear probability models. To make the interpretation less dependent on a particular functional form, we report additional robustness checks for the main empirical patterns. First, we verify that the treatment manipulation shifts first- and second-order beliefs in the intended direction. Second, we examine whether the treatment affects sharing conditional on the realized endowment. Third, we assess whether dictators' second-order beliefs are systematically associated with sharing behavior.

Throughout this section, the dependent variables are expressed in percentage points when using linear probability models. Standard errors in Table 3 are clustered at the participant level. Tables 4–7 use session-clustered standard errors as the primary inference. Appendix Tables A3–A4 use participant-clustered standard errors within the gender-specific subsamples, whereas the pooled tests of gender differences use session-clustered standard errors. We also report specifications with standard errors clustered at the session level, non-parametric tests, and logit/probit specifications with average marginal effects. Because beliefs are elicited on a five-point scale, we further replicate the regressions treating SOBs as an ordinal variable rather than as a cardinal standardized measure by using an ordered probit.

Let T_i^H indicate assignment to the treatment with a high probability of playing the x_L -Game: $T_i^H = 1$ for t_1 and $T_i^H = 0$ for t_2 . In the data, this corresponds to the treatment in which first- and second-order beliefs are higher on average. The variables FOB and SOB denote, respectively, recipients' first-order beliefs and dictators' second-order beliefs. We also construct a standardized measure of second-order beliefs, denoted SOB_i^z , and an ordinal measure, denoted SOB_i^o , increasing in the recipient's expected probability of sharing.

We first test whether the treatment shifts beliefs. Because the endowment is

realized only after beliefs are elicited, the manipulation-check regressions do not include the realized-endowment indicator. The baseline specifications are:

$$FOB_i = \alpha + \beta T_i^H + \delta R_i + \varepsilon_i, \quad (12)$$

$$SOB_i = \alpha + \beta T_i^H + \delta R_i + \varepsilon_i. \quad (13)$$

where R_i denotes the standardized round control.

Table 4 reports manipulation checks for the effect of the high-expectations treatment on recipients' first-order beliefs and dictators' second-order beliefs. Rows (1)–(4) report linear regressions where the dependent variable is expressed in percentage points. The reported coefficient is the effect of the high-expectations treatment. All regression specifications include a standardized round control; none includes the realized endowment. Because the treatment is common to all participants in a session-round, session-clustered standard errors in rows (2), (4), and (5) provide the primary inference. Rows (1) and (3) report participant-level clustering for comparison. Row (5) reports an ordered probit specification using the ordinal SOB measure, increasing in expected sharing.

Table 4: Manipulation checks

	coef.	s.e. rob.	t/z	p-value	95% CI	
	(a)	(b)	(c)	(d)	(e)	(f)
(1) FOB	13.11	1.48	8.87	< 0.001	10.21	16.02
(2) FOB, ses. clust.	13.11	1.54	8.51	< 0.001	9.82	16.40
(3) SOB	12.33	1.52	8.11	< 0.001	9.34	15.32
(4) SOB, ses. clust.	12.33	1.74	7.07	< 0.001	8.61	16.05
(5) Ordinal SOB, ses. clust.	0.395	0.059	6.67	< 0.001	0.279	0.510

Note: Treatment status varies at the session-round level, but the two counterbalanced treatment sequences were randomly assigned to sessions. Rows (2), (4), and (5) cluster standard errors at the session level (16 sessions) and provide the primary inference. Rows (1) and (3) use participant-level clustering for comparison. For the ordered probit model in row (5), column (c) reports the z-statistic rather than the t-statistic.

The treatment has a large and statistically significant effect on both first- and second-order beliefs. Recipients' first-order beliefs are about 13.1 percentage points higher in the high-expectations treatment, and dictators' second-order beliefs are about 12.3 percentage points higher. Both effects remain highly significant when standard errors are clustered at the session level, the primary level of inference. Individual-level clustering and an ordered probit model using the ordinal SOB measure confirm the same pattern.

We next examine whether the treatment affects sharing conditional on the realized endowment. This is the key reduced-form comparison for the belief-dependent channel because it holds fixed the dictator's realized material incentives. Let M_i be an indicator equal to one when the dictator is in the 40-Game and equal to zero in the 24-Game. For the high-endowment condition, we estimate:

$$share_i = \alpha + \beta T_i^H + \delta R_i + \varepsilon_i, \quad \text{for } M_i = 1. \quad (14)$$

where $share_i = 1$ when i chooses to share, otherwise it is zero.

Table 5 reports treatment effects on sharing conditional on the realized endowment. The reported coefficient is the effect of the high-expectations treatment. The primary specifications use mean sharing in each session-round-endowment cell and cluster standard errors at the session level. Table A1 in the Appendix reports individual-level LPM, logit, and probit sensitivity checks. Sharing and average marginal effects are expressed in percentage points throughout. Within Table 5, row (1), corresponding to $H2a$, is the sole primary test. Row (2), corresponding to $H2b$, and row (3), comparing the treatment effects across endowments, are secondary tests.

Table 5: Treatment effects on sharing by endowment

	coef.	s.e. rob.	t/z	p-value	95% CI	
(1) 40-Game	9.64	4.88	1.97	0.067	-0.77	20.04
(2) 24-Game	5.08	3.09	1.64	0.121	-1.51	11.66
(3) Difference: 40 vs. 24	4.56	5.10	0.89	0.386	-6.32	15.43

Note: All specifications include a control for the round in which the decision was made. Rows (1)–(3) use means for each session-round-endowment cell; there are 128 cells for each endowment and 16 sessions. Standard errors are clustered at the session level. Row (3) reports the interaction between treatment and the 40-Game indicator, i.e., the difference between treatment effects across endowments. Exact randomization inference enumerates all $\binom{16}{8} = 12,870$ balanced assignments of the two treatment sequences to sessions. The two-sided exact p -values corresponding to rows (1)–(3) are 0.071, 0.099, and 0.399, respectively. If the three tests are instead treated as a single family, Holm’s step-down correction yields adjusted two-sided exact p -values of 0.213, 0.213, and 0.399, respectively. Individual-level sensitivity checks are reported in Table A1.

In the primary session-round-cell analysis, sharing in the 40-Game is 9.64 percentage points higher in the high-expectations treatment. The estimate is marginally significant at the 10% level ($p = 0.067$). In the 24-Game, the corresponding estimate is 5.08 percentage points and is not statistically significant ($p = 0.121$). The individual-level specifications reported in Table A1 produce the same qualitative pattern, with positive treatment effects in both endowment conditions. Realized endowment subcells contain either four or twelve dictators; the primary analysis therefore gives each session-round equal weight.

Exact randomization inference based on the random assignment of the two counterbalanced sequences to sessions yields two-sided p -values of 0.071 for the 40-Game effect and 0.099 for the 24-Game effect. The corresponding one-sided p -values in the predicted positive direction are 0.036 and 0.050. Because $H2a$ is the sole primary behavioral test, its two-sided exact p -value is our primary inferential reference. As a conservative sensitivity, treating all three tests in Table 5 as a single family and applying Holm’s correction yields no rejection at conventional significance levels, reinforcing our interpretation of the treatment evidence as suggestive rather than conclusive.

Importantly, the pooled session-round-cell interaction test does not reject equality of the two treatment effects. The estimated difference between the effect in the 40-Game and the effect in the 24-Game is 4.56 percentage points ($p = 0.386$). The

corresponding two-sided exact randomization p -value is 0.399. Individual-level interaction estimates likewise do not reject equality (Table A1). Thus, although the point estimate is larger in the 40-Game, the data do not establish that treatment effects differ across endowment conditions.

Overall, the conditional treatment effects are positive in both endowment conditions. The evidence for the high-endowment condition, which is the main behavioral test of the belief-dependent channel, is positive and marginally significant in the primary session-round-cell analysis. Therefore, we interpret it as suggestive reduced-form evidence consistent with the guilt-aversion channel, rather than as stand-alone conclusive evidence.

We also examine whether sharing differs between the 24-Game and the 40-Game within each treatment. We estimate fixed-effects models of the form:

$$share_{isr} = \alpha_i + \mu_{sr} + \beta M_{isr} + \varepsilon_{isr}, \quad (15)$$

separately by treatment, where α_i and μ_{sr} denote subject and session-round fixed effects.

Table 6 reports within-treatment endowment effects. The reported coefficient is the effect of being in the 40-Game relative to the 24-Game. In the fixed-effects linear probability models, the dependent variable is expressed in percentage points. Standard errors are clustered at the session level, allowing arbitrary dependence across subjects and rounds within each of the 16 independent sessions. Conditional logit estimates use within-subject variation in sharing and drop subjects with no within-subject variation in the dependent variable.

Table 6: Within-treatment endowment effects

	coef.	s.e. rob.	t/z	p-value	95% CI	
Subject & cell FE LPM	(a)	(b)	(c)	(d)	(e)	(f)
(1) High-exp. treat.	−0.78	4.07	−0.19	0.851	−9.46	7.90
(2) Low-exp. treat.	−5.86	5.68	−1.03	0.319	−17.97	6.26
Conditional logit	(a)	(b)	(c)	(d)	(e)	(f)
(3) High-exp. treat.	−0.03	0.15	−0.21	0.837	−0.33	0.27
(4) Low-exp. treat.	−0.35	0.23	−1.48	0.140	−0.80	0.11

Note: Linear probability models include subject and session-round fixed effects.

Conditional logit specifications include a round control. All standard errors are clustered at the session level (16 sessions). For the conditional logit specifications, column (c) reports the z-statistic rather than the t-statistic.

The fixed-effects estimates do not provide robust evidence that sharing is lower in the 40-Game than in the 24-Game. The estimated coefficient on the 40-Game indicator is negative in both treatments, as predicted, but is not statistically significant in the models with subject and session-round fixed effects. Conditional logit estimates clustered at the session level lead to the same conclusion. Hence, these additional checks suggest that the within-treatment cost effect should be interpreted cautiously in this dataset.

Finally, we examine the association between dictators' second-order beliefs and sharing. These regressions are not interpreted as stand-alone causal evidence, because individual SOBs may partly reflect projection or intended behavior. They are instead used as mechanism evidence: if the belief-dependent interpretation is correct, higher SOBs should be associated with higher sharing.

The baseline mechanism specification is:

$$share_i = \alpha + \beta T_i^H + \gamma M_i + \eta(T_i^H \times M_i) + \lambda SOB_i^z + \delta R_i + \varepsilon_i. \quad (16)$$

Table 7 reports the association between dictators' second-order beliefs and sharing. In the linear probability models, the dependent variable is a binary measure of the dictator's decision, equal to 100 if the share option was chosen and 0 otherwise.

Baseline specifications are supplemented by models with subject and session-round fixed effects. Logit and probit entries report average marginal effects, expressed in percentage points.

Table 7: Second-order beliefs and sharing

	coef./AME	s.e. rob.	t/z	p-value	95% CI	
LPM	(a)	(b)	(c)	(d)	(e)	(f)
(1) Stand. SOB	9.40	1.51	6.22	< 0.001	6.18	12.62
(2) Ordinal SOB	7.43	1.19	6.22	< 0.001	4.88	9.97
(3) SOB in 24-Game	9.07	2.13	4.26	0.001	4.53	13.61
(4) SOB in 40-Game	9.72	2.07	4.69	< 0.001	5.30	14.15
Subject and cell FE	(a)	(b)	(c)	(d)	(e)	(f)
(5) Stand. SOB	5.34	2.32	2.30	0.036	0.40	10.29
(6) Ordinal SOB	4.22	1.83	2.30	0.036	0.32	8.13
AME of stand. SOB	(a)	(b)	(c)	(d)	(e)	(f)
(7) Logit	9.20	1.43	6.44	< 0.001	6.40	11.99
(8) Probit	9.24	1.44	6.41	< 0.001	6.42	12.07

Note: Baseline and nonlinear specifications control for treatment, endowment, their interaction, and round. Rows (5)–(6) include subject and session-round fixed effects. All standard errors are clustered at the session level (16 sessions), allowing arbitrary dependence across subjects and rounds within a session. For logit and probit specifications, column (c) reports the z-statistic rather than the t-statistic.

A one-standard-deviation increase in SOBs is associated with an increase of about 9.4 percentage points in the probability of sharing. The result is highly statistically significant with session-clustered standard errors. The same pattern holds when SOBs are treated as an ordinal variable rather than as a cardinal standardized measure. With subject and session-round fixed effects, the association becomes smaller but remains positive and significant: 5.34 percentage points for standardized SOBs ($p = 0.036$) and 4.22 points for ordinal SOBs ($p = 0.036$).

Logit and probit specifications also yield similar average marginal effects, around 9.2 percentage points.

The positive association between SOBs and sharing is therefore robust to the use of cardinal and ordinal belief measures, nonlinear binary-choice models, and demanding subject and session-round fixed effects. This supports the interpretation that beliefs are related to sharing decisions in the direction predicted by guilt aversion. However, as emphasized above, this association should not be interpreted as sufficient by itself to identify guilt aversion, since individual beliefs may also reflect projection or anticipated own behavior.

Taken together, the robustness checks yield three main conclusions. First, the treatment strongly shifts both first- and second-order beliefs in the intended direction. Second, conditional treatment effects on sharing are positive, with the high-endowment comparison providing suggestive but not uniformly statistically significant evidence consistent with a belief-dependent channel. Third, dictators' second-order beliefs are strongly and robustly associated with sharing. These checks therefore do not change the logic of the design: the main reduced-form evidence remains the comparison across treatments conditional on the high-endowment realization, interpreted jointly with the manipulation checks and the belief-sharing association.

6 Identification, alternative interpretations, and limits of the design

Our interpretation rests on the joint reading of the manipulation check, reduced-form treatment comparisons, and individual-level belief-behavior associations. This logic follows the experimental literature on belief-dependent preferences, which emphasizes that testing guilt aversion requires variation in expectations that is not merely inferred from observed behavior (Vanberg, 2008; Khalmetski, 2016; Di Bartolomeo *et al.*, 2019).

The design is strongest when it compares sharing across treatments conditional on the realized endowment, especially in the 40-Game.

In that comparison, the material payoff consequences of sharing and keeping are fixed, while the commonly known distribution of endowments differs across

treatments and induces different expectations.

A first possible concern is that the treatment may affect more than beliefs. Changing the probability of the low-endowment game also changes the ex ante frequency with which participants expect the environment to be relatively favorable to sharing. This may shift the perceived social meaning of the game or the salience of the dictator’s opportunity cost, a concern related to the broader evidence that framing and institutional details can affect giving in dictator games (List, 2007; Bardsley, 2008; Krupka and Weber, 2013). For this reason, we interpret the treatment as a manipulation of beliefs induced by the distribution of hidden endowments, not as a pure direct manipulation of recipient expectations holding every other aspect of the environment fixed.

A second concern is projection. Dictators’ individual SOBs may partly reflect what they themselves intend to do rather than an independent assessment of recipients’ expectations. This concern is common in belief-elicitation studies and is one reason why correlations between beliefs and behavior are not sufficient, by themselves, to identify guilt aversion (Ellingsen *et al.*, 2010; Khalmetski *et al.*, 2015; Khalmetski, 2016). It does not invalidate the treatment-level evidence, but it implies that the individual-level relationship between SOBs and sharing should be read as mechanism-consistent evidence rather than as a stand-alone causal estimate.

A third concern is that behavior may reflect distributional preferences, moral-standard concerns, or experimenter demand rather than guilt aversion. Altruism, maximin preferences, inequity aversion, and other distributional motives can generate sharing without belief dependence (Fehr and Schmidt, 1999; Charness and Rabin, 2002; Engel, 2011). Moreover, Hauge (2016) shows that dictator giving may respond not only to recipients’ expectations but also to recipients’ moral standards about what dictators should give, so a treatment-induced shift in perceived norms could mimic, or interact with, a guilt-aversion channel. Similarly, Stringhi (2026) emphasizes that guilt aversion and inequity aversion can be difficult to distinguish from observed behavior alone, since both can rationalize prosocial choices but imply different comparative statics under suitable payoff manipulations. Experimenter demand may also make the equal split salient (Zizzo, 2010; de Quidt *et al.*, 2018). Our design addresses these concerns by separating the within-treatment

effect of the realized endowment from the across-treatment effect of beliefs. Still, the results should be interpreted as evidence consistent with a belief-dependent guilt-aversion channel operating alongside belief-independent motives, rather than as a structural decomposition that uniquely rules out all alternatives.

Finally, the design has limited scope. The game is one-shot, the subject pool is drawn from undergraduate students, and the endowment distribution has only two realizations. These features make the experiment transparent and useful for identification, but they also limit external validity, as emphasized by the broader literature on laboratory experiments and subject pools (Falk and Heckman, 2009; Henrich *et al.*, 2010). Future work could vary recipient expectations more directly, introduce richer distributions of endowments, or study repeated and strategically richer environments.

7 Exploratory gender heterogeneity

Gender heterogeneity was not part of our core theory-driven tests and is treated as exploratory. The sample includes 231 female and 281 male participants. In pooled individual fixed-effects models with session-clustered standard errors, we do not reject equality between female and male participants either in the baseline SOB slope ($p = 0.547$) or in the SOB–endowment interaction ($p = 0.109$). We therefore find no statistically reliable evidence that the association between beliefs and sharing differs by gender. Gender-specific estimates are reported in Appendix Tables A3–A4 for descriptive completeness, while the formal between-gender comparisons are reported in Appendix Table A5. Differences in statistical significance within the two subsamples should not be interpreted as evidence of a between-gender difference.

8 Conclusions

This paper studies how prosocial behavior is shaped by the interaction of two mechanisms emphasized in the social-preferences literature. The first is a belief-

independent social-norm channel: dictators may share because they dislike violating an equal-sharing norm, and this motive becomes less powerful when sharing is materially more costly. The second is a belief-dependent guilt-aversion channel: dictators may share because they do not want to disappoint recipients by falling short of what they believe recipients expect.

The experimental design generates variation in both dimensions. The realized endowment, observed only by the dictator, changes the material opportunity cost of sharing. The treatment changes the commonly known probability of the low-endowment game and thereby shifts beliefs. This structure allows us to separate three empirical objects: whether the manipulation moves beliefs, whether sharing differs across treatments conditional on the realized endowment, and whether individual second-order beliefs are associated with sharing.

The results show, first, that the manipulation works as intended. When the low-endowment game is more likely, both first-order and second-order beliefs about sharing are higher. Second, conditional on the high-endowment realization, sharing is higher in the high-expectations treatment. This is our primary behavioral test of the belief-dependent channel, because keeping in this condition leaves the recipient with zero and is therefore most likely to generate disappointment. We interpret this pattern as evidence consistent with guilt aversion. Third, the within-treatment estimates are directionally consistent with lower sharing when the opportunity cost of sharing is higher, but they are imprecisely estimated and not statistically significant. They therefore provide only suggestive evidence consistent with a cost-sensitive norm channel.

The individual-level regressions reinforce this interpretation. Dictators with higher second-order beliefs are more likely to share, and higher endowments reduce sharing. The estimated interaction between beliefs and endowment is positive, as predicted by the model, but not statistically significant in the pooled sample. We therefore interpret the evidence as supporting a joint role for belief-independent and belief-dependent motives, rather than as a sharp structural decomposition of the two.

Several qualifications are important. The treatment changes beliefs through the distribution of hidden endowments; it does not manipulate recipient expectations directly while holding all other features of the environment fixed. Individual

second-order beliefs may also partly reflect projection or intended behavior. For these reasons, our strongest claim is not that we uniquely identify guilt aversion, but that the evidence is consistent with a belief-dependent channel operating alongside a cost-sensitive norm channel. The gender analysis is exploratory and should be viewed as hypothesis-generating.

Future work could strengthen the design by manipulating recipient expectations more directly, introducing richer endowment distributions, collecting larger samples for heterogeneity analysis, and studying repeated or strategically richer environments. Despite these limits, the paper provides a theory-guided test showing that the distribution of hidden endowments can shape expectations and that these expectations are behaviorally relevant for sharing.

Appendix A

A1. Belief elicitation

Elicitation of first-order beliefs. Recipients make their guess by ticking one of the five points in Table A. The five response categories form an ordinal elicitation device. For the empirical analysis, they are coded as 1, 0.75, 0.50, 0.25, and 0; the belief measures reported in the main tables are averages of these coded responses. These empirical codes are distinct from the probability values used to calculate incentive payments. The payment schedule in Table A was derived from a quadratic scoring rule evaluated at probabilities 85%, 68%, 50%, 32%, and 15% (see Vanberg, 2008, p. 1472).

Table A – Incentives for the elicitation of first-order beliefs (tokens)

A will	choose <i>Share</i>			choose <i>Keep</i>	
	Certainly	Probably	Unsure	Probably	Certainly
Please tick your guess	O	O	O	O	O
Your earnings if A chooses					
<i>Share</i>	0.65	0.60	0.50	0.35	0.15
<i>Keep</i>	0.15	0.35	0.50	0.60	0.65

Elicitation of second-order beliefs. After recipients submit their first-order belief, dictators are asked to guess which of the five points in Table A was selected by their counterpart. Correct guesses are paid 0.50 tokens.

A2. Sensitivity analyses

Table A1 reports the individual-level counterparts to the primary session-round-cell estimates in Table 5. The LPM, logit, and probit specifications yield positive treatment effects in both endowment conditions. The interaction estimates do not reject equality of the two treatment effects. The logit and probit sensitivity checks cluster standard errors at the session level, consistent with the assignment of treatment sequences.

Table A1: Individual-level sensitivity checks for treatment effects on sharing by endowment

	coef.	s.e. rob.	t/z	p-value	95% CI	
	(a)	(b)	(c)	(d)	(e)	(f)
40-Game						
(S1) LPM	7.03	3.14	2.24	0.026	0.86	13.20
(S2) LPM, session clust.	7.03	5.13	1.37	0.191	−3.90	17.96
(S3) logit AME, ses. clust.	7.03	5.12	1.37	0.170	−3.01	17.07
(S4) probit AME, ses. clust.	7.02	5.12	1.37	0.170	−3.01	17.05
24-Game						
(S5) LPM	4.88	3.20	1.53	0.128	−1.41	11.17
(S6) LPM, session clust.	4.88	2.40	2.03	0.060	−0.24	10.00
(S7) logit AME, ses. clust.	4.88	2.40	2.04	0.042	0.18	9.59
(S8) probit AME, ses. clust.	4.88	2.40	2.04	0.042	0.18	9.58
Difference: 40 vs. 24						
(S9) Interaction	2.15	4.46	0.48	0.630	−6.61	10.90
(S10) Interaction, ses. clust.	2.15	5.69	0.38	0.711	−9.98	14.28

Note: All specifications use individual observations and include a control for the round in which the decision was made. Rows (S1), (S5), and (S9) use participant-clustered standard errors. Rows (S2)–(S4), (S6)–(S8), and (S10) use session-clustered standard errors (16 sessions). For logit and probit specifications, column (c) reports

the z-statistic rather than the t-statistic. Rows (S9)–(S10) report the interaction between treatment and the 40-Game indicator. Sharing and average marginal effects are expressed in percentage points.

Table A2 reports a sensitivity analysis for the individual fixed-effects model in Table 3. The coefficient estimates are unchanged, while standard errors are clustered at the session level to allow arbitrary dependence among participants and decisions within each of the 16 sessions. The SOB coefficient in the 24-Game remains positive and statistically significant (5.69, s.e. 2.26, $p = 0.024$). The implied SOB slope in the 40-Game is also positive and statistically significant (7.46, s.e. 2.06, $p = 0.002$).

Table A2: Table 3 sensitivity to session-level clustering

	coef.	s.e. rob.	t	p-value	95% CI	
	(a)	(b)	(c)	(d)	(e)	(f)
(1) SOBs (24-Game)	5.69	2.26	2.52	0.024	0.87	10.51
(2) endowment	−5.68	2.34	−2.43	0.028	−10.67	−0.69
(3) SOBs $\times$ endowment	1.77	2.50	0.71	0.491	−3.57	7.11
(4) round	−1.55	0.90	−1.72	0.105	−3.46	0.37
(5) SOBs (40-Game)	7.46	2.06	3.62	0.002	3.07	11.85

Note: The specification is identical to Table 3 and includes participant fixed effects and a standardized round control. Standard errors are clustered at the session level (16 sessions). The estimation sample contains 2,048 dictator-round observations from 512 participants. Row (5) reports the linear combination of the SOB coefficient and the SOB–endowment interaction. Coefficients are expressed in percentage points; tests and confidence intervals use 15 degrees of freedom.

A3. Exploratory gender heterogeneity

Motivated by evidence from experiments in related settings (Charness *et al.*, 2024; Di Bartolomeo *et al.*, 2024), Tables A3 and A4 report the Table 3 specification separately for female and male participants. These split-sample estimates are presented for descriptive completeness and do not provide a direct test of differences between genders.

Table A3: Gender-specific sharing regression: females

	coef.	s.e. rob.	t	p-value	95% CI	
	(a)	(b)	(c)	(d)	(e)	(f)
(1) SOBs	4.32	2.60	1.66	0.098	-0.81	9.44
(2) endowment	-2.81	3.15	-0.89	0.374	-9.03	3.40
(3) (1) $\times$ (2)	6.68	3.09	2.17	0.031	0.60	12.77
(4) round	-2.97	1.22	-2.43	0.016	-5.38	-0.56
(5) constant	43.41	1.57	27.70	< 0.001	40.32	46.49

Table A4: Gender-specific sharing regression: males

	coef.	s.e. rob.	t	p-value	95% CI	
	(a)	(b)	(c)	(d)	(e)	(f)
(1) SOBs	6.74	2.58	2.61	0.010	1.66	11.83
(2) endowment	-8.22	3.06	-2.69	0.008	-14.24	-2.21
(3) (1) $\times$ (2)	-2.01	3.21	-0.63	0.531	-8.33	4.30
(4) round	-0.34	1.38	-0.25	0.806	-3.05	2.37
(5) constant	44.51	1.54	28.96	< 0.001	41.48	47.53

Note: Both specifications are individual fixed-effects linear probability models with a standardized round control. Standard errors are clustered at the participant level within each subsample. The female and male samples contain 231 and 281 participants, respectively, each contributing four dictator decisions. Differences in statistical significance across the two tables are not tests of gender differences. Formal pooled between-gender comparisons are reported in Table A5.

Table A5: Pooled tests of exploratory gender heterogeneity

	coef.	s.e. rob.	t	p-value	95% CI	
	(a)	(b)	(c)	(d)	(e)	(f)
(1) SOBs	4.16	2.57	1.62	0.126	-1.32	9.64
(2) endowment	-2.88	3.54	-0.81	0.428	-10.43	4.66
(3) (1)×(2)	6.71	3.14	2.14	0.050	0.02	13.40
(4) Male × SOBs	2.60	4.22	0.62	0.547	-6.40	11.61
(5) Male × endowment	-5.13	5.25	-0.98	0.344	-16.33	6.07
(6) Male × (1)×(2)	-8.74	5.13	-1.70	0.109	-19.68	2.20
(7) round	-1.54	0.89	-1.72	0.105	-3.44	0.36
(8) constant	43.95	1.20	36.63	< 0.001	41.40	46.51

Note: The omitted categories are female participants and the 24-Game. The coefficient in row (4) tests the gender difference in the baseline SOB slope. The coefficient in row (6) tests whether the SOB-endowment interaction differs by gender. The specification is an individual fixed-effects linear probability model with a standardized round control. The time-invariant male indicator is absorbed by the participant fixed effects. Standard errors are clustered at the session level (16 sessions). The estimation sample contains 2,048 dictator-round observations from 512 participants; tests and confidence intervals use 15 degrees of freedom. Coefficients are expressed in percentage points.

A6. A sufficient condition for the Auxiliary Hypothesis

We present a condition about the primitives of the model that implies the Auxiliary Hypothesis, i.e., that — keeping beliefs fixed — the norm-channel for changes in endowments prevails, and thus the sharing rate is strictly higher when the endowment is low, x_L , rather than high, x_H .

Let F_n (resp. F_g) denote the cdf of θ_n (resp. θ_g) in the subpopulation of norm-driven (resp. guilt-driven) dictators, and recall that q_n denotes the fraction of norm-driven dictators.

For the purpose of deriving a sufficient condition, it is enough to assume that all guilt-driven dictators have the same SOB β . With this, the sharing rate with

endowment x and SOB β is

$$\text{Fr}(\text{Share}; \beta, x) = q_n \left(1 - F_n \left(\hat{\theta}_n(x) \right) \right) + (1 - q_n) \left(1 - F_g \left(\hat{\theta}_g(\beta, x) \right) \right). \quad (17)$$

Proposition 3 *The norm-channel prevails—in the sense that the increase in sharing among norm-driven dictators more than offsets the maximal possible decrease in sharing among guilt-driven dictators—and therefore the aggregate sharing rate is higher under the low endowment if*

$$q_n \left(F_n \left(x_H - \frac{M}{2} \right) - F_n \left(x_L - \frac{M}{2} \right) \right) > (1 - q_n) (1 - F_g(1)).$$

Proof. Recall that the indifference threshold for a norm-driven dictator with endowment $x \in \{x_L, x_H\}$ is

$$\hat{\theta}_n(x) = x - \frac{M}{2}.$$

Given that in the parametrization of our experiment a dictator with high endowment is certain to disappoint the recipient, the indifference threshold for a guilt-driven dictator with *high endowment* is

$$\hat{\theta}_g(\beta, x_H) = \frac{x_H - \frac{M}{2}}{\bar{m}(\beta) - (M - x_H)} = \frac{M/2}{\bar{m}(\beta)} \geq 1,$$

because $x_H = M$ and $\bar{m}(\beta) := \mathbb{E}_\beta(\mathbb{E}_\alpha(\mathbf{m}_R)) \in [0, M/2]$. Note that

$$\inf_{\beta} \hat{\theta}_g(\beta, x_H) = 1 \quad (18)$$

is attained when $\bar{m}(\beta) = M/2$, i.e., when the dictator is certain that the recipient is certain that the dictator will share.

When the endowment is *low*, the expected disappointment is also low, and the dictator has little or no incentive to share. In particular,

$$\sup_{\beta} \hat{\theta}_g(\beta, x_L) = +\infty \quad (19)$$

is attained for every SOB β such that $\mathbb{P}_\beta(\mathbb{E}_\alpha(\mathbf{m}_R) \leq M - x_L) = 1$.

Therefore,

$$\text{Fr}(\text{Share}; \beta, x_L) > \text{Fr}(\text{Share}; \beta, x_H)$$

if and only if

$$\begin{aligned} & q_n \left(1 - F_n \left(x_L - \frac{M}{2} \right) \right) + (1 - q_n) \left(1 - F_g \left(\hat{\theta}_g(\beta, x_L) \right) \right) \\ > & q_n \left(1 - F_n \left(x_H - \frac{M}{2} \right) \right) + (1 - q_n) \left(1 - F_g \left(\hat{\theta}_g(\beta, x_H) \right) \right). \end{aligned}$$

Since

$$\begin{aligned} & \text{Fr}(\text{Share}; \beta, x_L) - \text{Fr}(\text{Share}; \beta, x_H) \\ & \geq \inf_{\beta} (\text{Fr}(\text{Share}; \beta, x_L) - \text{Fr}(\text{Share}; \beta, x_H)) \\ & \geq \inf_{\beta} \text{Fr}(\text{Share}; \beta, x_L) - \sup_{\beta} \text{Fr}(\text{Share}; \beta, x_H), \end{aligned}$$

eq.s (17), (18), and (19) imply that sharing is higher under low endowment if

$$q_n \left(1 - F_n \left(x_L - \frac{M}{2} \right) \right) > q_n \left(1 - F_n \left(x_H - \frac{M}{2} \right) \right) + (1 - q_n) (1 - F_g(1)),$$

that is, if

$$q_n \left(F_n \left(x_H - \frac{M}{2} \right) - F_n \left(x_L - \frac{M}{2} \right) \right) > (1 - q_n) (1 - F_g(1)). \blacksquare$$

References

- Attanasi, G., P. Battigalli, and E. Manzoni (2016), “Incomplete information models of guilt aversion in the trust game,” *Management Science*, 62, 648-667.
- Attanasi, G., P. Battigalli, E. Manzoni, and R. Nagel (2025), “Disclosure of belief-dependent preferences in a trust game,” *Economic Theory*, 10.1007/s00199-025-01645-5.
- Bartling, B., F. Engl, and R. A. Weber (2014), “Does willful ignorance deflect punishment? An experimental study,” *European Economic Review*, 70: 512-524.

- Battigalli, P., and M. Dufwenberg (2007), “Guilt in games,” *American Economic Review*, 97: 170-176.
- Battigalli, P., and M. Dufwenberg (2022), “Belief-dependent motivations and psychological game theory,” *Journal of Economic Literature*, 60: 833-882.
- Battigalli P., R. Corrao, and M. Dufwenberg (2019): “Incorporating belief-dependent motivation in games,” *Journal of Economic Behavior and Organization*, 167, 185-218.
- Battigalli, P., G. Di Bartolomeo, and S. Papa (2025), “Guilt aversion and inequality in dictator games,” *Economics Letters*, 247, February.
- Bardsley, N. (2008), "Dictator game giving: Altruism or artefact?," *Experimental Economics*, 11, 122–133.
- Benabou, R., and J. Tirole (2006), “Incentives and prosocial behavior,” *American Economic Review*, 96(5): 1652-1678.
- Cartwright, E. (2019), “A survey of belief-based guilt aversion in trust and dictator games,” *Journal of Economic Behavior and Organization*, 167: 430-444.
- Charness, G., and M. Rabin (2002), "Understanding social preferences with simple tests," *Quarterly Journal of Economics*, 117, 817–869.
- Charness, G., G. Di Bartolomeo, and S. Papa (2024), “A stranger in a strange land: Promises and identity,” *Games and Economic Behavior*, 144, 13-28.
- Chowdhury, S. M., and J. Y. Jeon (2014), “Impure altruism or inequality aversion? An experimental investigation based on income effects,” *Journal of Public Economics*, 118(C): 143-150.
- Ciccarone, G., G. Di Bartolomeo, and S. Papa (2020), “The rationale of in-group favoritism: An experimental test of three explanations.” *Games and Economic Behavior*, 124, 554-568.
- Cox, J. C. (2004), “How to identify trust and reciprocity,” *Games and Economic Behavior*, 46: 260-281.
- Cox, J. C., V. Sadiraj, and U. Schmidt (2015), “Paradoxes and mechanisms for choice under risk.” *Experimental Economics*, 18, 215-250.
- d’Adda, G., M. Dufwenberg, F. Passarelli and G. Tabellini (2020), “Social norms with private values: Theory and experiments,” *Games and Economic Behavior*, 124: 288-304.

- Danilov, A., K. Khalmetski, and D. Sliwka (2021), "Descriptive norms and guilt aversion," *Journal of Economic Behavior and Organization*, 191: 293–311.
- de Quidt, J., J. Haushofer, and C. Roth (2018), "Measuring and bounding experimenter demand," *American Economic Review*, 108, 3266–3302.
- Di Bartolomeo, G., and S. Papa (2016), "Trust and reciprocity: Extensions and robustness of triadic design," *Experimental Economics*, 19: 100–115.
- Di Bartolomeo, G., M. Dufwenberg, S. Papa, and F. Passarelli (2019), "Promises, expectations & causation," *Games and Economic Behavior*, 113: 137–146.
- Di Bartolomeo, G., M. Dufwenberg, and S. Papa (2023a), "Promises and partner-switch," *Journal of the Economic Science Association*, 9, 77–89.
- Di Bartolomeo, G., M. Dufwenberg, and S. Papa (2023b), "The sound of silence: A license to be selfish," *Economics Letters*, 182, 68–70.
- Di Bartolomeo, G., M. Dufwenberg, S. Papa, and L. Razzolini (2024), "Guilt aversion and moral commitment: Eve versus Adam," *Journal of the Economic Science Association*, 1–17.
- Ellingsen, T., M. Johannesson, S. Tjøtta, and G. Torsvik (2010), "Testing guilt aversion," *Games and Economic Behavior*, 68(1): 95–107.
- Engel, C. (2011), "Dictator games: A meta study," *Experimental Economics*, 14, 583–610.
- Falk, A., and J. J. Heckman (2009), "Lab experiments are a major source of knowledge in the social sciences," *Science*, 326: 535–538.
- Fehr, E., and K. M. Schmidt (1999), "A theory of fairness, competition, and cooperation," *The Quarterly Journal of Economics*, 114(3): 817–868.
- Fehr, E., and K. M. Schmidt (2010), "On inequity aversion: A reply to Binmore and Shaked," *Journal of Economic Behavior and Organization*, 73: 101–108.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178.
- Gächter, S., L. Gerhards, and D. Nosenzo (2017), "The importance of peers for compliance with norms of fair sharing," *European Economic Review*, 97: 72–86.
- Khalmetski, K. (2016), "Testing guilt aversion with exogenous shift in beliefs," *Games and Economic Behavior*, 97: 110–119.

- Hauge, K. E. (2016), "Generosity and guilt: The role of beliefs and moral standards of others," *Journal of Economic Psychology*, 54, 35-43.
- Henrich, J., S. J. Heine, and A. Norenzayan (2010), "The weirdest people in the world?," *Behavioral and Brain Sciences*, 33, 61-83.
- Khalmetski, K., A. Ockenfels, and P. Werner (2015), "Surprising gifts: Theory and laboratory evidence," *Journal of Economic Theory*, 159: 163-208.
- Korenok, O., E. L. Millner, and L. Razzolini (2012), "Are dictators averse to inequality?," *Journal of Economic Behavior and Organization*, 82(2-3): 543-547.
- Korenok, O., E. L. Millner, and L. Razzolini (2014), "Taking, giving, and impure altruism in dictator games," *Experimental Economics*, 17(3): 488-500.
- Krupka, E. L., and R. A. Weber (2013), "Identifying social norms using coordination games: Why does dictator game sharing vary?," *Journal of the European Economic Association*, 11(3): 495-524.
- List, J. A. (2007), "On the interpretation of giving in dictator games," *Journal of Political Economy*, 115, 482-493.
- Murnighan, J. K., J. M. Oesch, and M. Pillutla (2001), "Player types and self-impression management in dictator games: Two experiments," *Games and Economic Behavior*, 37: 388-414.
- Nguyen, C. V. (2019), "The effect of inequality in stakes on sharing behavior: Evidence from an experimental study," *Economics Letters* 184, 108638.
- Servátka, M. (2009), "Separating reputation, social influence, and identification effects in a dictator game," *European Economic Review*, 53: 197-209.
- Stringhi, A. (2026), "Guilt and fairness," *Games and Economic Behavior*, 158, 191-218.
- Vanberg, C. (2008), "Why do people keep their promises? An experimental test of two explanations," *Econometrica*, 76: 1467-1480.
- Zizzo, D. J. (2010), "Experimenter demand effects in economic experiments," *Experimental Economics*, 13, 75-98.