Experimental Economics (2024) 27:491–547
https://doi.org/10.1007/s10683-024-09827-z

Avoiding the cost of your conscience: belief dependent preferences and information acquisition

Claire Rimbaud[1] · Alice Soldà[2]

Received: 31 August 2021 / Revised: 25 March 2024 / Accepted: 14 April 2024 /
Published online: 27 May 2024

Abstract

Pro-social individuals typically face a trade-off between their monetary incentives and their other-regarding preferences. When this is the case, they may be tempted to exploit the uncertainty in their decision environment to reconcile this trade-off. In this paper, we investigate whether individuals with belief-dependent preferences acquire information about others' expectations in a self-serving way. We present a model of endogenous information acquisition and test our theoretical predictions in an online experiment based on a modified trust-game in which the trustee is uncertain about the trustor's expectations. Our experimental design enables us to (1) identify participants with belief-based preferences and (2) investigate their information acquisition strategy. Consistent with our predictions for *subjective* belief-dependent preferences, we find that most individuals classified as belief-dependent strategically select their source of information to avoid the cost of their conscience.

Keywords Belief-dependent preferences · Information acquisition · Self-serving biases · Experiment

JEL Classification C91

The replication and supplementary material for the study is available at https://osf.io/fujdg/?view_only=0ec4ab35410e477997765ddb7549c34e.

✉ Claire Rimbaud
claire.rimbaud@dauphine.psl.eu

Alice Soldà
alice.solda@ugent.be

[1] Université Paris Dauphine-PSL, LEDa (UMR CNRS 8007), 75016 Paris, France

[2] Department of Economics, RISLαβ, Ghent University, Sint-Pietersplein 6, 9000 Ghent, Belgium

1 Introduction

A large body of evidence has shown that individuals often care about the welfare of others.[1] These pro-social individuals typically face a trade-off between their monetary incentives and their other-regarding preferences, and might be tempted to exploit the uncertainty in their decision environment to reduce the tension between the two. In a seminal paper, later replicated by Larson and Capra (2009) and Feiler (2014), Dana et al. (2007) exposed this trade-off by showing that individuals behave more selfishly when they are uncertain about the consequences of their choice on others' payoffs.[2] The fact that people use the uncertainty in their environment as an excuse for their selfish behavior has been supported by subsequent research. For instance, individuals have been shown to manipulate their beliefs about others' intentions (Di Tella et al., 2015; Andreoni & Sanchez, 2020) or to take advantage of the uncertainty on whether their choices would be implemented (Haisley & Weber, 2010; Exley, 2016; Garcia et al., 2020) to behave more selfishly.

This growing body of evidence has focused on *outcome*-based preferences, i.e., preferences over the allocation of payoffs between oneself and others. Yet, pro-social behavior can also be shaped by *belief*-based preferences, i.e., preferences over the allocation of payoffs between oneself and others *that are conditional on beliefs*.[3] For example, let's imagine that Ann hires Bob to work on a job for her in exchange for a fixed wage. Ann holds private expectations about how much Bob should work on the job, given how much she pays him. If Bob is purely selfish, he maximizes his utility function by providing zero effort, regardless of his beliefs about Ann's expectations. In contrast, Bob's preferences regarding his level of effort may be sensitive to Ann's expectations. For instance, Bob may experience guilt from disappointing Ann's expectations and feel compelled to provide a high level of effort if he believes Ann expects him to do so. If Bob doesn't know exactly Ann's expectations, he may be tempted to use this uncertainty to maintain the belief that Ann does not expect much from him so as to provide little effort without feeling guilty.

We investigate whether individuals with belief-dependent preferences engage in self-serving information acquisition when they are uncertain about others' expectations. To do so, we examine the information acquisition strategy of decision-makers with belief-dependent preferences who face a conflict between their monetary interest and their other-regarding preferences in the context of a trust game. We first

[1] For instance, people donate positive amounts of money to others without any strategic incentives to do so (Forsythe et al., 1994) or prefer more equitable monetary allocations over selfish ones (Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000). People donate more when their recipient expects to receive more (Bellemare et al., 2018; Attanasi et al., 2019), or lie less often when others can infer their degree of dishonesty (Dufwenberg and Dufwenberg, 2018).

[2] Serra-Garcia and Szech (2022) show that this demand for less information is sensitive to the cost of avoiding the information.

[3] *Outcome*-based preferences include, for instance, altruism and spitefulness (Levine, 1998), inequality-aversion (Fehr & Schmidt, 1999) or ERC-preference (Bolton & Ockenfels, 2000). While *belief*-based preferences include, for instance, regret theory (Loomes & Sugden, 1982), guilt-aversion (Battigalli & Dufwenberg, 2007) or expectation-based reciprocity (Dufwenberg & Kirchsteiger, 2004).

present a theoretical framework adapted from the model of endogenous information acquisition proposed by Spiekermann and Weiss (2016). In our framework, the second-mover ('trustee') is uncertain about the first-mover's ('trustor') expectations and can acquire information to resolve this uncertainty. We distinguish between trustees with 'subjective' preferences (i.e., preferences that depend on what trustees *believe* about the trustor's expectations) and trustees with 'objective' preferences (i.e., preferences that depend on the trustor's *actual* expectations). Within this framework, we demonstrate that it is optimal for trustees with subjective belief-dependent preferences to bias their information acquisition strategy towards signals that reduce the tension between their monetary payoff and their other-regarding preferences.[4]

We then conducted an online experiment in which we can classify trustees as either belief-dependent or belief-independent by observing their decisions in the trust game. As in the theoretical framework, trustees were initially uncertain about the trustors' expectations and were later provided with an opportunity to acquire information about these expectations. Crucially, trustees faced two information sources that were skewed in opposite directions. This feature of the design allows us to assess whether belief-dependent trustees biased their information search in a way that is congruent with their monetary incentives.

We find that 44% of trustees in our sample can be classified as belief-dependent. Among these belief-dependent trustees, 60.47% strategically acquire signals that lead to higher expected payoffs (i.e., lower amounts sent back to the trustor), which is consistent with our predictions for subjective preferences. Our findings highlight yet another channel through which people make selfish choices while keeping a 'good conscience'.

Our main contribution is to the literature on *strategic* information acquisition. While there is extensive evidence that individuals can deliberately remain ignorant about the consequences of their actions (see Golman et al., 2017; Hertwig & Engel, 2016 for reviews),[5] a small body of research has now shown that individuals can also actively seek information if, in expected terms, selfish justifications become more available by doing so. When the information acquisition choice is binary (acquiring the information or not), Fong and Oberholzer-Gee (2011) find that dictators who choose to acquire information about why their recipient is 'poor' use it as an excuse to reduce their donations. When information is acquired sequentially, individuals stop collecting information earlier when they liked early returns (Ditto & Lopez, 1992; Smith et al., 2017; Chen et al., 2021). We differentiate ourselves from these lines of research by focusing on situations in which individuals can discriminate between sources of information. In this literature, individuals have been shown to prefer positively skewed, confirmatory, or less informative information sources

[4] Other models based on different mechanisms can also predict strategic information acquisition by assuming belief-dependent preference, such as the model of moral constraints by Rabin (1995) self-signaling theories (e.g., Grossman & Van Der Weele, 2017) or a model relying on an aversion to harm others (Chen et al., 2021).

[5] In particular, Xiao and Bicchieri (2012) shown that information avoidance can be relevant in the context of empirical expectations. The authors found that, when there is a small cost to acquire information, dictators avoid information on injunctive or descriptive norms about giving.

(Spiekermann & Weiss, 2016; Soraperra et al., 2023; Soldà et al., 2020; Charness et al., 2021; Chopra et al., 2023). The current paper especially relates to Spiekermann and Weiss (2016), who allow participants to strategically seek and/or avoid information about the descriptive norms of donations. We add to this emergent literature by showing that individuals can also strategically discriminate between more or less self-serving information sources when information relates to others' expectations.[6]

We also contribute to a recent strand of papers calling attention to the impact of situational excuses on the expression of belief-dependent preferences (e.g., Balafoutas & Fornwagner, 2017; Inderst et al., 2019; Morell, 2019). In that respect, we are the first to show that people try to remain uncertain about others' *beliefs* to make selfish choices while appearing as if they cared about others. Two studies also studied how the uncertainty about others' *intentions* could lead to the formation of self-serving beliefs (Di Tella et al., 2015; Friedrichsen et al., 2022).[7] Di Tella et al. (2015) study a corruption game where dictators are uncertain on whether their recipient is making a 'side deal' to obtain a larger benefit from the allocation of tokens. When recipients have the option to make such a deal, dictators are more selfish than when this deal is made randomly by a computer (i.e., without selfish intentions). The authors conclude that dictators distort their beliefs about the recipients' intentions to justify their selfish allocations. While they investigate how uncertainty affects participants' beliefs and decisions, we move one step further by studying whether belief-dependent participants adopt self-serving strategies to resolve the uncertainty.

The remainder of the paper is organized as follows. We develop our theoretical model in Sect. 2. In Sect. 3, we present the experimental design used to address our research question. Next, we derive our experimental hypotheses in Sect. 4. In Sect. 5, we describe our empirical results. Finally, we conclude in Sect. 7.

2 Theoretical model

We introduce a modified trust game with incomplete information. The first mover ("trustor") decides between two actions: *In* or *Out*. If the trustor chooses *In*, the second mover ("trustee") receives an endowment E to allocate between himself and the trustor.[8] The trustee returns an amount y (with $0 \leq y \leq E$) to the trustor and keeps $E - y$ to himself. If the trustor chooses *Out*, the game ends, and

[6] It also broadly relates to the recent work by Saccardo and Serra-Garcia (2023) which investigates whether people who can choose a decision environment more favorable to self-serving decisions eventually act self-servingly (even though they *actively* chose to be in this favorable environment). Our question falls under the same umbrella as we ask whether people who claim to care about others' expectations *actively* acquire information about these expectations in a self-serving manner. Importantly, we are able to answer this question within-subjects.

[7] In a recent working paper, Jia (2021) also finds that personal experience can impact people's empathy, where empathy is defined as one's beliefs about others' feelings.

[8] For the sake of clarity, we use "she/her" when referring to the trustor and "he/him" when referring to the trustee.

each player receives an outside option. The trustee receives $O^{trustee}$, and the trustor receives $O^{trustor}_{\omega}$, which depends on the state of the world $\omega \in \{L(ow), H(igh)\}$, with $O^{trustor}_{H} > O^{trustor}_{L}$. The trustor knows her outside option with certainty when choosing between *In* or *Out*. In contrast, the trustee does not know the trustor's outside option when choosing y, but knows that both outside options are equally likely, that is $p = p(\omega = L) = 1 - p(\omega = L) = 0.5$. Importantly, the trustee can acquire costless signals about the trustor's outside option before choosing how much to return (details in Sect. 2.2). The structure of the game is summarized in Fig. 1.

We define $\phi_{\omega} \in [0, E]$, the trustor's expectations about his payoff conditional on choosing *In*, where $\phi_{\omega} = \mathbb{E}^{trustor}[y|In, \omega]$; and $\Phi_{\omega} \in [0, E]$, the trustee's beliefs about the trustor's expectations: $\Phi_{\omega} = \mathbb{E}^{trustee}[\phi_{\omega}]$. We refer to the former as the trustor's first-order beliefs and to the latter as the trustee's second-order beliefs.

2.1 Belief formation

A payoff-maximizing trustor will choose *In* only if she expects to receive more from doing so than from choosing *Out*, i.e., when Eq. 1 is satisfied.

$$\phi_{\omega} \geq O^{trustor}_{\omega} \tag{1}$$

Assuming that the distribution of trustors' first-order beliefs has mass between the two possible values of their outside options, a payoff-maximizing trustor will hold higher beliefs when her outside option is High than when her outside option is Low, conditional on choosing *In*: $\phi_L \leq \phi_H$. Using psychological forward induction reasoning (Dufwenberg, 2002), a trustee will be able to infer Eq. 1.[9] More precisely, the trustee understands that a trustor will only choose *In* if she expects to receive at least her outside option by doing so. Hence, the trustee's second-order beliefs also increase in the trustor's outside option: $\Phi_L \leq \Phi_H$. It leads to the following assumption.

Assumption Conditional on choosing In, trustors' first-order beliefs and trustees' second-order beliefs are higher when the outside option is High rather than Low.

Note that we make the implicit assumption that the trustor's outside option only affects trustees' behavior through these second-order beliefs. We discuss the implication of this assumption in more details in Sect. 6.

2.2 Belief-dependent preferences

The core of our analysis focuses on individuals with belief-dependent preferences. A belief-dependent trustee's utility function (Eq. 2) depends on his material payoff, $E - y$, and his belief-dependent motivation, $c(y, \phi_{\omega})$. The belief-dependent

[9] Experimental evidence in favor of the psychological forward induction reasoning was provided by Woods and Servátka (2016).

motivation (Eq. 3) is the absolute difference between how much the trustor expects to receive (ϕ_ω) and how much the trustor actually receives (y). This psychological component of the utility function is weighted by the trustee's sensitivity to his belief-dependent motivation, denoted γ_i (Eq. 3), which can be positive or negative. When $\gamma_i > 0$, the trustee experiences a psychological cost from returning an amount y that deviates from the trustor's expectations. This cost can stem from various psychological considerations such as cognitive dissonance (Festinger, 1962), a distaste for violating social norms (Spiekermann & Weiss, 2016), an aversion to inflict reliance damage (Sengupta & Vanberg, 2023), or guilt-aversion (Battigalli & Dufwenberg, 2007). When $\gamma_i < 0$, the trustee experiences a psychological gain from deviating from the trustor's expectations. This behavior, while less intuitive at first sight, is also consistent with a wide range of psychological traits including a preference to surprise others (Khalmetski et al., 2015) or a preference to reward 'benevolence', which relates to the concept of expectation-based reciprocity (Dufwenberg & Kirchsteiger, 2004).[10]

$$u_i(y, \phi_\omega) = (E - y) - c(y, \phi_\omega) \tag{2}$$

$$\text{with } c(y, \phi_\omega) = \gamma_i \cdot |\phi_\omega - y| \tag{3}$$

The optimal amount y^* the trustee returns to the trustor depends on the trustee's sensitivity to his belief-dependent motives: When $\gamma_i \in]-1, 1[$, the trustee assigns a higher weight to his monetary payoff than to his belief-dependent motivation. Therefore, he will behave as a payoff-maximizer and return $y^* = 0$.[11] When $\gamma_i \notin]-1, 1[$, the trustee assigns more value to his belief-dependent motivation than to his monetary payoff. In other words, he is '*sufficiently* belief-dependent' so that his choices could be distinguishable from those of a payoff-maximizer. This second case splits into two: When $\gamma_i > 1$, the trustee experiences a psychological cost from deviating from the trustor's expectations, and $u_i(y, \phi_\omega)$ is maximized for $y^* = \phi_\omega$. We refer to such trustees as 'belief-concordant'. When $\gamma_i < -1$, the trustee experiences a psychological gain from deviating from the trustor's expectations and y^* depends on the level of ϕ_ω. When $\phi > \frac{E}{2}$, $u_i(y, \phi_\omega)$ is maximized for $y^* = 0$. When $\phi < \frac{E}{2}$, $u_i(y, \phi_\omega)$ is maximized for $y^* = 0$ when $\gamma > \frac{E}{E-2\phi}$, and $y^* = E$ otherwise. We refer to such trustees as 'belief-discordant'. Proposition 1 below summarizes the optimal amount returned y^* depending on γ_i and ϕ. The proofs are provided in the Appendix A.2.

Proposition 1 *When* $\gamma_i \in]-1, 1[$, *belief-dependent trustees return* $y^* = 0$. *When* $\gamma_i > 1$, *belief-dependent trustees return* $y^* = \Phi_\omega$. *When* $\gamma_i < -1$, *belief-dependent trustees return either* $y^* = 0$ *or* $y^* = E$ *depending on the level of* ϕ_ω.

[10] While we are now agnostic regarding the underlying motives that would lead a trustee to exhibit either 'belief-concordant' or 'belief-discordant' preferences, an earlier version of the paper attempted to motivate the research by drawing only on guilt-aversion and expectation-based reciprocity. However, as pointed out by the editor, our design does not allow a precise identification of these types as we cannot perfectly rule out that a positive correlation between y^* and ϕ_ω stems from other forms of reciprocity.

[11] If $\gamma_i = 1$ and $\gamma_i = -1$, then $u_{c,i} = E - \phi_\omega$, hence the solution is indeterminate: $y^* \in [0, E]$.

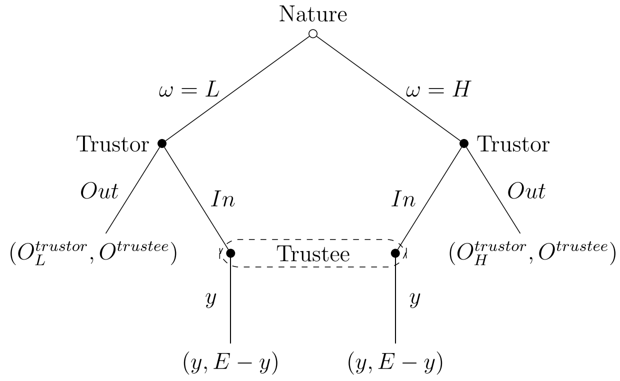


Fig. 1 Trust game with High or Low outside option

In the remainder of this section, we restrict our analysis to the case where the trustee faces a trade-off between his monetary payoff and his belief-dependent motivations (i.e., sufficiently belief-dependent trustee). Since very few trustees can be classified as belief-discordant in our experiment and both cases are symmetric, we provide the details of the analysis for belief-concordant trustees in the main text below and relegate the corresponding analysis for belief-discordant to Appendix A.1.

Decision Under Certainty. Let $\hat{u}_i(\phi_\omega) = \max_y u_i(y, \phi_\omega)$ be the maximum utility achievable for a belief-concordant trustee for a given expectation of the trustor ϕ_ω. This function decreases with ϕ_ω: the higher the expectations of the trustor, the less the trustee keeps for himself. The proof is provided in Sect. A.3. Recalling our auxiliary assumption, which states that $\phi_L < \phi_H$, it follows that $\hat{u}_i(\phi_L) > \hat{u}_i(\phi_H)$. When the trustor's expectations are low, a belief-concordant trustee reaches the maximum utility $\hat{u}_i(\phi_L)$ by returning $y_L^* = \phi_L$. When the trustor's expectations are high, the maximum utility $\hat{u}_i(\phi_H)$ is reached by returning $y_H^* = \phi_H$.

Decision Under Uncertainty. We now turn to the situation where the trustee is initially uncertain about the trustor's expectation *ex-ante* and can acquire costless signals about the trustor's expectation before choosing how much to return. We define p' as the updated probability that the trustor's expectation is Low after the acquisition of signal(s). The trustee can acquire one or two types of signals, represented by the random variables S_L and S_H. With probability s, the signal S_ω reveals the true expectation of the trustor, that is, the signal reveals that the trustor's expectation is Low if the trustor's expectation is indeed Low ($S_\omega = L$); or that the trustor's expectation is High if the trustor's expectation is indeed High ($S_\omega = H$); and with probability $1 - s$, the signal does not reveal the trustor's expectation (null signal, $S_\omega = 0$). After a null signal, the trustee updates the probability p using Bayes' rule, which yields a posterior of $p'_{ns} = \frac{(1-s)p}{(1-s)p+(1-p)}$ after $S_L = 0$ and $p'_{ns} = \frac{p}{p+(1-s)(1-p)}$ after $S_H = 0$. Finally, if the trustee receives both $S_L = 0$ and $S_H = 0$, no update is necessary as the two signals cancel out each other: $p'_{ns} = p$.

Objective vs. Subjective preferences. To analyze the information acquisition strategy of a belief-concordant trustee, we distinguish between *objective* and *subjective* preferences. For a belief-concordant trustee with *objective* preferences, the psychological cost from a mismatch depends on what the trustor's true expectation ϕ_ω actually is. Therefore, his psychological component is a function of ϕ_ω which can take two values, either ϕ_L when the trustor's expectation is Low or ϕ_H when the trustor's expectation is High. In contrast, for a belief-concordant trustee with *subjective* preferences, the psychological cost from a mismatch depends on *his second-order belief* Φ about the trustor's expectation, and not on the trustor's actual expectation. Crucially, under uncertainty, Φ can differ from ϕ_ω.

Objective Preferences Under Uncertainty. Under uncertainty, a belief-concordant trustee with *objective* preferences cannot be sure to choose the action that minimizes his psychological cost and must instead minimize the expected psychological cost given by: $p \cdot c(y_U, \phi_L) + (1-p) \cdot c(y_U, \phi_H)$. Therefore, the maximum expected utility under uncertainty for a given p is achieved when the trustee returns the amount $y^*_U(p)$. Now recall that a belief-concordant trustee with objective belief-dependent preferences maximizes his utility when his return matches the trustor's actual expectation. This implies that $\hat{u}_{c,i}(y^*_H, \phi_H) > \hat{u}_{c,i}(y^*_U(p), \phi_H)$ and $\hat{u}_{c,i}(y^*_L, \phi_L) > \hat{u}_{c,i}(y^*_U(p), \phi_L)$. Hence, the information acquisition strategy that maximizes his utility is to acquire both signals, as it maximizes his chances to learn about the trustor's actual expectation and therefore return the optimal amount y^*_ω. The proof is provided in Sect. A.4.

Proposition 2 *Objective belief-concordant trustees acquire both signals.*

Subjective Preferences Under Uncertainty. For a belief-concordant trustee with *subjective* preferences, his psychological cost from a mismatch depends on the epistemic state of the world Φ_p. We follow Spiekermann and Weiss (2016) in proposing a coarse mapping from states to beliefs defined as the step function Φ_p (Eq. 4) characterized by the probability p that the state is Low.[12] As in Spiekermann and Weiss (2016), we consider three epistemic states: knowing that the trustor's actual expectation is Low ($\Phi_p = \Phi_L$), knowing that the trustor's actual expectation is High ($\Phi_p = \Phi_H$), or not knowing the trustor's actual expectation ($\Phi_p = \Phi_U$). In Eq. 4, the parameter $\epsilon \in [0, \frac{1}{2}[$ represents the degree of 'caution' with which a subjective trustee interprets any probability p. As ϵ tends to 1/2, the trustee treats any probability p greater than 1/2 *as if* $p = 1$, and any probability lower than 1/2 *as if* $p = 0$. Note that, when $\epsilon = 1/2$, the state Φ_U disappears. Symmetrically, as ϵ tends to zero, the trustee becomes more 'cautious' in his interpretation of p. Crucially, we are making the assumption that a null signal never removes uncertainty to ensure that any update from p to p'_{ns} does not change ϕ_p. It implies $\epsilon < p'_{ns}$.

[12] In the context of compliance to social norms, Spiekermann and Weiss (2016, p. 174) argue that 'since degrees of beliefs are not observable in detail, it is unlikely that social norms take them as argument with any great precision'. We consider that the same reasoning applies to belief-dependent preferences.

$$\Phi_p = \begin{cases} \Phi_L & \text{if } p \geq 1-\epsilon \\ \Phi_U & \text{if } \epsilon < p < 1-\epsilon \\ \Phi_H & \text{if } p \leq \epsilon \end{cases} \tag{4}$$

Unlike belief-concordant trustees with objective preferences, a belief-concordant trustee with subjective preferences conditions the amount to be returned on Φ_p and not ϕ_ω. Recall that the maximum utility achievable is decreasing in the trustor's expectations. This implies that $\hat{u}_i(y^*_L, \Phi_L) > \hat{u}_i(y^*_U, \Phi_U) > \hat{u}_i(y^*_H, \Phi_H)$. Consequently, a belief-concordant trustee with subjective preferences will sample information from the signal S_L only, as it increases the probability of receiving the highest utility $\hat{u}_i(\Phi_L)$ without any down-side risk.[13] The proof is provided in Sect. A.5.

Proposition 3 *Subjective belief-concordant trustees acquire a Low signal only.*[14]

To summarize, belief-concordant trustees would prefer to be in the state where the trustor's expectation is low so as to return little, irrespective of whether their preferences are objective or subjective. However, a belief-concordant trustee with objective preferences cares about the trustor's actual expectation ϕ_ω and is therefore better off with more information as he cannot change the state he is in. In contrast, a belief-concordant trustees with subjective preferences only cares about his beliefs Φ_p about the trustor's expectation and therefore has an incentive to strategically acquire information that maximizes his chances to learn that the trustor's expectations are low.

2.3 Belief-independent preferences

Some trustees may return the same amount y irrespective of their belief about the trustor's expectations. For instance, a pure payoff-maximizing trustee will always return zero, and an inequality-averse trustee will return the same positive amount (at most $\frac{E}{2}$) regardless of his beliefs about the trustor's expectations. Because our research question focuses on belief-dependent preferences, we pool these preference types together under the label 'belief-independent' preferences. Our theoretical model is agnostic regarding what belief-independent trustees should do. This is because the information is payoff-irrelevant to them, as (by definition) belief-independent trustees return the same amount irrespective of their beliefs about the trustor's expectations. Hence, they have no incentives to systematically favor one information source over the other.

[13] Note that this result stems from the assumption of a coarse mapping from states to beliefs. If we were to assume a linear mapping instead, it is straightforward that the optimal choice of a trustee with subjective belief-dependent preferences would be to avoid information altogether.

[14] Note that Proposition 3 holds because we make the assumption that a null signal never removes uncertainty. Otherwise, a belief-concordant trustee with subjective preferences would prefer to avoid information altogether.

3 Design

To test our theoretical predictions, we designed an experiment that mimics the theoretical framework described in Sect. 2. Within this framework, we first introduced uncertainty about the trustors' expectations and then provided trustees with an opportunity to acquire information to alleviate this uncertainty.

3.1 Experiment outline

Trust game. At the beginning of the experiment, participants are randomly allocated to either the role of a trustor or trustee. A trustor faces two options: *Out* and *In*. If she chooses *Out*, the game ends, and both players receive their respective outside option. The trustees' outside option is equal to 90 cents.[15] In contrast, the trustors' outside option depends on the game being played. If a trustor chooses *In*, she foregoes her outside option. As a consequence, the trustee receives 200 cents to allocate between himself and his trustor in increments of 15 cents. Both the trustor and the trustee are informed of the entire payoff structure, including the existence of two equally likely outside options for the trustor. However, trustors are informed about their outside option before they make their decision, while trustees do not know which of the two outside options the trustor is facing at the time of decision.

Outside option manipulation. The trustor's outside option depends on the game being played. In the Low game, trustors receive 15 cents if they choose *Out*. In contrast, trustors receive 75 cents if they choose *Out* in the High game. This feature of the design creates an exogenous variation in the participant's beliefs about the trustor's expected payoffs from choosing *In*. We operate under the assumption that trustors who choose *In* expect a return at least equal to the outside option that they were willing to forego. Therefore, conditional on choosing *In*, (i) trustors' first-order beliefs about their own payoff should be higher when the outside option is High rather than Low, and (ii) anticipating this, trustees' second-order beliefs about the trustors' payoff should also be higher when the outside option is High rather than Low.

Beliefs elicitation. Before trustors learn whether they are playing the Low game or the High game, we elicit their conditional beliefs about their expected payoffs from choosing *In* using the strategy method. More specifically, we ask trustors to indicate how much they expect to receive from their trustee if they choose *In* in the Low game and how much they expect to receive in the High game. Trustors' beliefs corresponding to the true state of the world are then matched with their trustee's decision. If a trustor's belief is accurate, with a 15 cents margin of error, he or she is paid 50 cents. Using a similar method, we also asked trustees to indicate how much they believed their trustor expected to receive if they chose *In*, both in the Low game and the High game. Trustees' beliefs corresponding to the true state of the

[15] All amounts are in USD.

world are then matched with their trustor's belief in that state. Trustees receive 50 cents if their beliefs are accurate with a 15 cents margin of error.

Trustee's return choices. Because trustees do not know their trustor's actual outside option at the time of their decision, we elicit how much trustees want to send to the trustor both (i) in case they *learn* that the outside option is Low (Decision Low) and (ii) in case they *learn* that the outside option is High (Decision High).[16] Trustees are informed that if they learn that the trustor's actual outside option is Low, Decision Low is implemented. Symmetrically, if they learn that the trustor's actual option is High, Decision High is implemented. If they remain uninformed about their trustor's actual outside option, the trustor receives the average of Decision Low and Decision High. This key feature of the design, inspired by the 'menu' method of Bellemare et al. (2011), is crucial to identify trustees with belief-dependent preferences.[17] Indeed, because trustors' outside options are designed to induce a shift in beliefs, eliciting trustees' returns conditional on their knowledge of the different outside options is equivalent to eliciting their choices conditional on the trustors' first-order beliefs.

Trustee's information acquisition. After making their conditional transfer decisions, trustees are unexpectedly offered the opportunity to acquire information about their trustor's outside option.[18] We use an information structure similar to Spiekermann and Weiss (2016): Trustees face four envelopes of two different colors: silver and gold. Trustees know that if their trustor's outside option is Low, the information is in one of the two silver envelopes, and the three other envelopes are empty. In contrast, if their trustor's outside option is High, the information is in one of the two gold envelopes, and the three other envelopes are empty. Hence, a silver envelope can never reveal that the trustor's outside option is high, and a gold envelope can never reveal that the trustor's outside option is low. Trustees can choose to open either (i) one silver envelope, (ii) one gold envelope, or (iii) one silver and one gold. The order of presentation of the envelopes was randomized at the participant level. While opening one envelope from each color maximizes the chances to learn the trustor's actual outside option, a trustee can strategically bias his information acquisition by opening a single envelope only. After selecting which envelope they wish to open, trustees are informed about the content of the selected envelopes, and the

[16] Note that because trustees did not observe the trustor's decision, they were asked to make a decision in the eventuality that their trustor chose *In* (strategy method).

[17] As mentioned in Sect. 2, this identification strategy relies on the assumption that the trustor's outside option affects trustees' conditional return decisions solely via their beliefs about the trustor's expectations (see Sect. 6).

[18] While trustees are informed that they may discover the trustor's outside option and how it will affect their payoff before making their conditional transfer decisions, they do not know that they will be able to actively acquire information.

computer program automatically implements the corresponding transfer decision.[19] The information acquisition procedure is summarized in Fig. 2.

Post-experimental questionnaires. Because our results rely on trustees' ability to infer their trustor's expectations from their trustor's potential outside options, participants' level of reasoning may affect their responsiveness to our treatment manipulation. Therefore, we elicit participants' level of reasoning using a 2/3 beauty-contest game in which participants are asked to indicate a number between 0 and 100, and are rewarded with 100 cents if the number they indicate corresponds to two-thirds of the mean of the numbers indicated by all participants enrolled in the experiment. Participants were also asked to report their age, gender, employment status, annual income, and weekly expenditure. Finally, we asked participants to rate the clarity of the instructions using a scale from 'extremely unclear' to 'extremely clear'. In addition, we asked trustees to explain their information acquisition decision in a free-form format, and participants were rewarded 50 cents to provide an answer.

3.2 Experimental strategy

Our experimental strategy relies on two conditions: First, we must be able to observe the signal choices of trustees with belief-dependent preferences. Second, a coarse-grained rule for conditional return under uncertainty must be induced successfully. In this section, we explain our use of the 'menu method' to classify trustees according to their preference type and our method to create or reinforce a coarse-grained mapping of beliefs into return decisions.

'Menu' Method. To classify trustees as either belief-dependent or belief-independent, we need to identify whether trustees' return decision varies with the trustor's expectations at the individual level. Consequently, we cannot rely on a between-subject design to address our main research question. Instead, we use a variant of the strategy method (the 'menu method') which allows us to separate trustees with incentives to strategically acquire information and trustees without.[20]

Coarse-grained mapping. The restriction of the trustees' action space under uncertainty allows us to implement the theoretical assumption of a coarse-grained mapping of belief into action. With the amount of uncertainty fixed externally, trustees cannot adjust their return decision as a response to a change in their belief about the trustor's expectations in a linear fashion. Theoretically, the amount returned under uncertainty, y_U, can take any value between y_L and y_H. We chose to impose $y_U = \frac{y_L+y_H}{2}$ as it corresponds to the amount that a trustee ex-ante expects to return

[19] Note that because trustees' final transfer depends directly on what they learn about the trustor's outside option in this stage, information directly impacts the players' payoff in our setting. This feature of the design is crucial to study participants' information acquisition strategies *given* their preferences. This differs from the original paradigm introduced by Dana et al. (2007) where the authors study participants' preferences *given* their choice of information.

[20] One might be worried that eliciting participants' conditional preferences might generate an experimenter demand effect. We discuss this concern in Sect. 6 and provide evidence from a follow-up study showing that it is unlikely to play a role in our experiment.

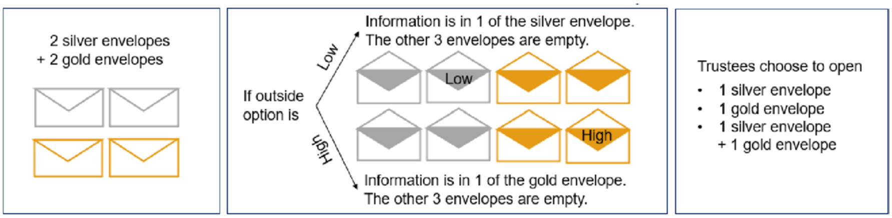

2 silver envelopes
+ 2 gold envelopes
If outside option is
Low
High
Information is in 1 of the silver envelope. The other 3 envelopes are empty.
Low
High
Information is in 1 of the gold envelope. The other 3 envelopes are empty.
Trustees choose to open
• 1 silver envelope
• 1 gold envelope
• 1 silver envelope + 1 gold envelope

Fig. 2 Choosing a source of information

(given that each state is ex-ante equally likely), and it has the advantage that it can be easily explained to the participants without having to introduce probabilities. It also ensures that the monetary incentives of acquiring either signal are the same in absolute terms for all trustees.

Procedures. We conducted the experiment online on Amazon MTurk. We recruited a total of 320 participants from the United States of America.[21] Participation was restricted to individuals over 18 years old who completed at least 300 HITs with an approval rate of at least 99%. Participants were randomly allocated the role of trustor or trustee at the beginning of the experiment. Pairs were formed after all participants had completed the experiment. During the experiment, participants could re-read the instructions at any time by clicking on a reminder button at the top of their screen.[22] Moreover, they had to answer a comprehension questionnaire correctly after the presentation of the instructions to proceed to the next step. Participants were paid less than 48 h after the completion of the experiment.

4 Experimental hypotheses

Our main research question is to assess whether individuals who exhibit belief-dependent preferences acquire information in a self-serving way. We address this question in two steps. First, we classify trustees as either belief-dependent or belief-independent.[23] Second, we test whether the information acquisition strategy of trustees we identified as either belief-concordant or belief-discordant differs from the information acquisition strategy of trustees we identified as belief-independent.

In Sect. 2, we assume that a higher trustor's outside option increases both trustors' first-order beliefs and trustees' second-order belief about the trustor's payoff

[21] With that sample size, the minimum detectable effect size with statistical power at the recommended.80 level (Cohen, 2013) is 0.44 for comparisons of the proportion of each information acquisition strategy between belief-independent and belief-dependent trustees, which is sufficient to detect an effect of half the magnitude of the one observed in Spiekermann and Weiss (2016).

[22] The screens used in the experiment are provided in Appendix D.

[23] Note that, we are only able to identify trustees whose preferences are *sufficiently* belief-dependent as defined in Sect. 2.2.

from choosing *In*. This assumption can hold in theory under certain conditions (see Sect. 2.1) but empirically it remains to be tested.

Auxiliary Hypothesis 1 Conditional on choosing In, trustors' first-order beliefs and trustees' second-order beliefs increase with the trustor's outside option.

If Auxiliary Hypothesis 1 is verified, we can classify trustees based on the correlation between their beliefs about the trustor's expectations and their conditional return decisions. Following Dufwenberg et al. (2011), we classify trustees with a positive correlation profile as 'belief-concordant' (i.e., return increases with beliefs about the trustor's expectations) and trustees with a negative correlation profile as 'belief-discordant' (i.e., return decreases with beliefs about the trustor's expectations). Finally, we classify trustees as belief-independent if their conditional return decisions are the same irrespective of their beliefs about the trustor's expectations. This leads to our second auxiliary hypothesis.

Auxiliary Hypothesis 2 The proportion of trustees identified as having belief-dependent preferences is strictly positive.

Our main research question is to identify whether trustees who exhibit belief-dependent preferences acquire information in a self-serving way. Opening both a silver and a gold envelope maximizes a trustee's chances to learn the trustor's actual outside option. However, some trustees with subjective preferences might be tempted to bias their information acquisition strategy towards signals that minimize the tension between their monetary incentives and their other-regarding motives. A subjective belief-concordant trustee keeps more money for himself when the trustor's outside option is Low. Consequently, a *subjective* belief-concordant trustee should open the silver envelope only as doing so allows him to learn that the trustor's outside option is Low if it is actually Low, while avoiding learning that the trustor's outside option is High if it is actually High. Symmetrically, an expectation-based reciprocal trustee keeps more money for himself when the trustor's outside option is High. Therefore, a *subjective* belief-discordant trustee should only open a gold envelope as doing so allows him to learn that trustor's outside option is High if it is actually High, while avoiding learning that the trustor's outside option is Low if it is actually Low. Finally, belief-independent trustees have no incentive to systematically favor one information source over the other (as information is payoff-irrelevant to them). Therefore, they provide a reasonable benchmark for comparing the information acquisition strategy of belief-dependent trustees. This leads to our two main hypotheses.

Hypothesis 1 belief-concordant trustees are more likely to open the silver envelope only compared to belief-independent trustees.

Hypothesis 2 belief-discordant trustees are more likely to open the gold envelope only compared to belief-independent trustees.

These hypotheses were pre-registered on AsPredicted.[24]

5 Experimental results

In this section, we first evaluate whether beliefs are affected by the outside option manipulation. We then classify trustees according to their type of preferences: belief-concordant, belief-discordant, or belief-independent. Finally, we assess how trustees' preferences affect their information acquisition strategy.

5.1 Are beliefs affected by the outside option manipulation?

In this section, we assess whether Auxiliary Hypothesis 1 is verified, that is, whether trustors' first-order beliefs and trustees' second-order beliefs about trustors' payoff from choosing *In* are higher in the High game than in the Low game.

Figure 3 displays the combination of beliefs about trustors' expected payoffs from choosing *In* in the Low game (x-axis) and in the High game (y-axis). The left panel displays trustors' beliefs, while the right panel displays trustees' beliefs. Figure 3 shows that there is a lot of heterogeneity in participants' responsiveness to the outside option manipulation.

The majority of participants' beliefs verify Auxiliary Hypothesis 1. We find that 53.13% of trustors, and 61.25% of trustees held higher beliefs in the High game than in the Low game (i.e., observations above the 45-degree line). In contrast, 10% of trustors and 8.13% of trustees indicated higher beliefs in the Low game than the high Game (i.e., observations below the 45-degree line), and 28.75% of trustors and 38.75% of trustees indicated similar expectations regardless of the game being played (i.e., observations on the 45-degree line). Interestingly, there seems to be a strong focal point around the egalitarian allocation, with 50% of the participants holding undifferentiated beliefs indicating beliefs at 90 cents in both games. To test our theoretical predictions, the subsequent analyses focus on the sub-sample of trustees who satisfied Auxiliary Hypothesis 1.

Result 1 53.13% of trustors' first-order beliefs and 61.25% of trustees' second-order beliefs are higher when the outside option is High rather than Low.

5.2 Are trustees motivated by belief-dependent preferences?

In this section, we classify trustees who satisfy Auxiliary Hypothesis 1 as belief-concordant, belief-discordant or belief-independent based on their conditional

[24] Link: https://aspredicted.org/blind.php?x=9md4uc. Note that the pre-registered hypotheses refer to 'guilt-averse' and 'expectation-based reciprocal' instead of 'belief-concordant' and 'belief-discordant' trustees. As mentioned earlier in Sect. 2, our experimental design does not allow us to cleanly identify these two types and we have adjusted the terminology in the paper accordingly. However, the substance of the hypothesized relationship remains unchanged.

transfers. Figure 4 displays the combinations of trustees' returns in the Low game (x-axis) and the High game (y-axis). We classify trustees who returned more in the High than in the Low game as belief-concordant (i.e., observations above the 45-degree line) and trustees who returned more in the Low than in the High game as belief-discordant (i.e., observations below the 45-degree line). Finally, trustees who returned the same amount regardless of the game are classified as belief-independent (i.e., observations on the 45-degree line).

About half of the trustees can be classified as belief-independent (52.04%, n = 51), returning on average 50.00 cents (se = 6.14). This average return hides two focal points where the trustees' payoff is maximized (returning 0 cents), and where equality is maximized (returning 90 cents). We found that 43.88% (n = 43) of trustees can be classified as belief-concordant. The average amount returned by belief-concordant trustees is 87.91 cents (se = 3.37) in the High game and 53.37 cents (se = 5.02) in the Low game. Only 4 trustees can be classified as belief-concordant.[25]

Result 2 There is a positive proportion of trustees exhibiting belief-dependent preferences in our sample: 43.88% of trustees can be classified as belief-concordant and 4.08% of trustees can be classified as belief-discordant.

5.3 How do belief-based preferences affect information acquisition?

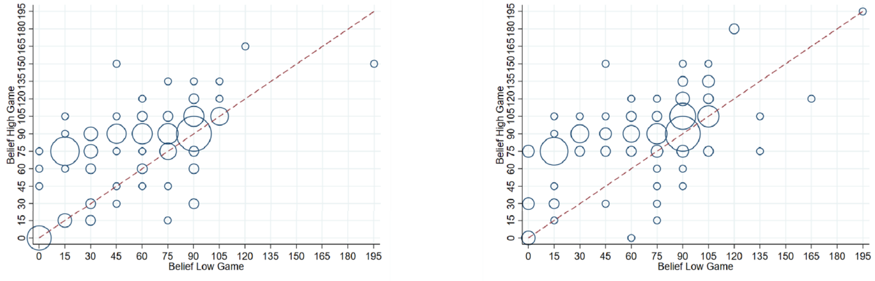
We now examine whether belief-independent and belief-dependent trustees adopt different information acquisition strategies. Given that only four trustees can be classified as belief-discordant, we restrict our analysis of information acquisition to trustees that were classified as either belief-independent or belief-concordant.

Figure 5 displays the distribution of information acquisition strategies for belief-independent (left-hand side) and belief-concordant trustees (right-hand side). It shows that the majority of belief-independent trustees chose to open both envelopes (52.94%), and they did so significantly more frequently than they would by chance (binomial test, $H0 = 0.33$, $p = 0.004$). We found that 17.65% of belief-independent trustees opened a silver envelope only, and 29.41% opened a gold envelope only. These results suggest that the default choice in the absence of strategic concerns is to acquire as much information as possible.[26] The post-experimental questionnaire allows us to investigate potential explanations for the trustees' information acquisition strategy. It revealed that 75% of belief-independent trustees who chose to open both envelopes indicated that they did so out of curiosity.[27]

[25] This is consistent with Attanasi et al. (2022) who find that positive correlation profiles are predominant in interactions in which the decision-maker's choice set is determined by a first-mover's willingness to blindly trust him/her.

[26] Belief-independent trustees earn the same payoff irrespective of what they learn, as their conditional returns are the same regardless of the trustor's outside option.

[27] This result is based on the answers of the 43 out of 51 belief-independent trustees who did provide an answer. The distribution of answers can be found in Table 3 in the Appendix.



(a) Trustors' first-order beliefs

(b) Trustees' second-order beliefs

Fig. 3 Distribution of individual beliefs about trustors' expected payoff from *In*

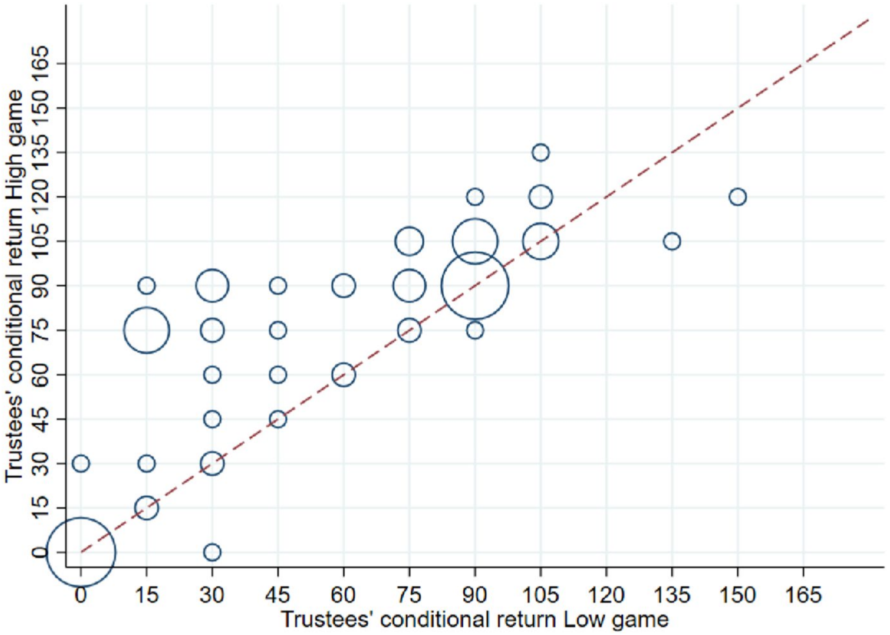


Fig. 4 Trustees' return strategies

In contrast to belief-independent trustees, the majority of belief-concordant trustees chose to open a silver envelope only (60.47%), and they did so significantly more frequently than they would by chance (binomial test, $H0 = 0.33$, $p < 0.001$). In addition, the proportion of belief-concordant trustees who chose to open a silver envelope only is significantly higher than the proportion of belief-independent trustees who made the same choice (Pearson's chi-square test, $p < 0.001$). The results of multinomial logit regressions reported in Table 4 in the Appendix corroborate these findings. Altogether, these observations suggest that the majority

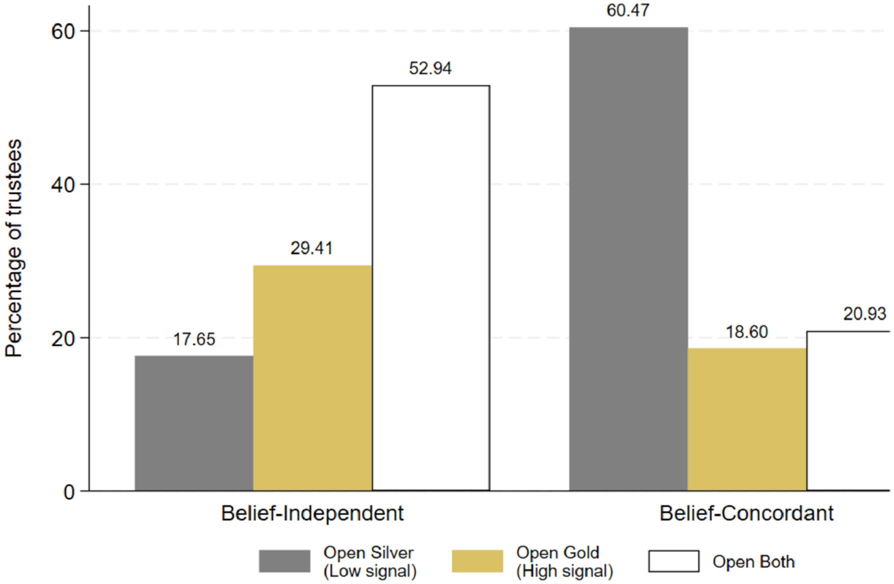


Fig. 5 Distribution of information acquisition strategies for belief-independent and belief-concordant trustees

of belief-concordant trustees exhibit an information acquisition strategy consistent with subjective preferences. Moreover, such strategy is indeed self-serving, as trustors receive significantly less in expectations (two-sided t-test: $p < 0.001$) when matched with a belief-concordant trustee who chooses to open the silver envelope only (66.32, $s.e. = 4.23$) than with a belief-concordant trustee who chooses to open both envelopes (70.64, $s.e. = 4.00$).

Result 3 belief-concordant trustees are more likely to open only the silver envelope compared to belief-independent trustees.

Although our model is agnostic on the relative proportions of the different information acquisition strategies, it is noteworthy that 20.93% of belief-concordant trustees chose to open both envelopes, which is consistent with our prediction for objective preferences. In contrast 18.60% of belief-concordant trustees chose to open only the gold envelope.[28] This information acquisition strategy cannot be explained by our model. We contend that it is likely due to behavioral noise. Indeed, this share goes down to 5.56% when excluding trustees who reported that (i) they did not understand that their choice of envelopes was payoff-relevant (n=6) or (ii) the instructions were not "extremely clear" (n=24) (see, Sect. C.2 in the Appendix).

[28] Both of these proportions are significantly lower than the proportion of trustees opening a silver envelope only (Wilcoxon signed rank tests: $p = 0.004$ and $p = 0.002$, respectively).

5.4 Determinants of trustees' information acquisition

We show in Sect. 5.3 that most belief-concordant trustees exhibit an information acquisition strategy consistent with subjective preferences. Hence, one can wonder whether these trustees are the ones that benefit the most from doing so. Indeed, one can imagine that trustees with the most money to lose from learning about a specific state of the world (i.e., trustees with more differentiated return decisions) are the most likely to engage in self-serving information acquisition strategies.

To address this question, we estimate multinomial logit models in which the dependent variable is a categorical variable that summarizes the three possible information acquisition strategies. The main explanatory variable corresponds to the difference between the amount returned by a belief-concordant trustee when he/she learns that the trustor's outside option is high and the amount returned when he/she learns that the trustor's outside option is low. The average marginal effects (AME) are displayed in Table 1.

Results from columns (1) and (2) show that a 10-cent increase in the difference in conditional returns leads to a 3% increase in the likelihood to open the silver envelope only.[29] However, the results are not significant. This null result goes against the idea that trustees with the most money to lose from learning about a specific state of the world are the most likely to engage in self-serving information acquisition strategies.

6 Discussion

We now examine the robustness of our findings in light of our choices regarding the identification strategy, design, and procedures. We discuss potential concerns and report additional analyses from our main experiment, as well as results from a follow-up experiment in an attempt to mitigate them.

Identification strategy. To classify trustees as either belief-dependent or belief-independent, we rely on the assumption that the change in the trustor's outside options affects the trustee's decision only indirectly via his beliefs. However, it is possible that our manipulation of the trustor's outside option directly affects trustees' behavior. For instance, trustees may care about the sacrifice the trustor makes by choosing *In* and derive utility from rewarding this sacrifice independently of the trustor's expectations.[30] Trustees with such preferences would return more when the outside option is High for reasons that are unrelated to their second-order beliefs. To disentangle these two channels, we turn to trustees who reported the same beliefs regardless of the trustor's outside option, as these trustees' return decisions cannot be driven by their beliefs. Comparing the behavior of such trustees to the behavior

[29] Considering a modal belief-concordant trustee who returns 15 cents in the low game and 75 cents in the high game, this coefficient implies that such trustee's probability to open the silver envelope is between 18 and 24% higher than for a belief-independent trustee.

[30] This can relate to informal concepts of reciprocity illustrated in McCabe et al. (2003) or the concept of 'reliance damage' by Sengupta and Vanberg (2023).

of trustees who reported higher second-order beliefs in the High game, we find no suggestive evidence in our data for this alternative channel. Indeed, among trustees who reported no difference in their second-order beliefs, 15% returned more when the outside option was higher (i.e., higher sacrifice) while this percentage rises to 44% among trustees who reported higher second-order beliefs when the outside option was higher. This finding reinforces our confidence in the validity of the main assumption underlying our identification strategy.

Experimenter demand effect. While the menu method has been widely used to elicit belief-dependent preferences in the literature (e.g., Khalmetski, 2016; Hauge, 2016; Bellemare et al., 2017; Bellemare et al., 2018), one might be concerned that the use of the strategy method might induce an experimenter demand effect (Zizzo, 2010).[31] Because both decisions are elicited on the same screen, participants may feel compelled to provide a different answer for each elicitation, which may not reflect their true preferences. If this is the case, we may be overestimating the role that belief-dependent preferences play in explaining participants' information acquisition strategy. We believe that demand effects should be limited in our context. Indeed, Bellemare et al. (2017) compare how different elicitation methods affect participants' responses in the trust game and found that the 'menu' method yields results similar to the "baseline" approach. In addition, the sizeable share of participants exhibiting belief-independent preferences also suggests that participants did not feel compelled to condition their return decisions on the trustor's outside option.

Nevertheless, we attempted to mitigate this concern by conducting a follow-up experiment similar to the original experiment described in Sect. 3 to the exception of three important changes (see the new decisions screens in Sect. D.3). First, we elicited trustees' transfer decisions conditional on the trustor's outside option being low and conditional on the trustor's outside option being high on two separate screens, and the order of the decisions screens was randomized at the participant level. Second, we did not remind trustees about how much money the trustor would give up by choosing *In* based on the trustor's outside option to make the trustors' expectations less salient. Finally, we did not remind trustees of their beliefs about the trustor's expectations to avoid priming participants into thinking that their beliefs about the trustor's expectations should matter.[32] We conducted the follow-up experiment on Amazon MTurk where we recruited 320 participants using the same procedure as in the original experiment. The results of the follow-up experiment are consistent with the results from the main experiment. Notably, we found that the share of participants conditioning the amount of money to return to the trustor on the trustor's outside option was much higher in the follow-up experiment than in the original one (65.93% vs. 43.88%). If anything, it seems that eliciting decisions on the same

[31] We thank an anonymous reviewer for highlighting this concern.

[32] A between-subjects design is typically used to limit demand effects in experiments. However, our research question relies on our ability to identify belief-dependent trustees, which requires us to elicit at least two data points per trustee.

Table 1 Average marginal effects of monetary incentives on the likelihood of each sampling strategy

	Open Silver		Open Gold		Open both	
	(1)	(2)	(3)	(4)	(5)	(6)
Return(High)–Return(Low)	0.003	0.004	– 0.000	0.000	– 0.003	– 0.004
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Ind. controls	No	Yes	No	Yes	No	Yes
Observations	43	43	43	43	43	43

Table reports the average marginal effects of our multinomial logit model of the difference in conditional returns on the likelihood of a given sampling strategy. Controls include the amount guessed in the beauty contest game and socio-demographic characteristics (age, identifying as a female, annual income, weekly expenditure). The sample is restricted to belief-concordant trustees only. Standard errors are in parentheses

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

screen leads to more consistency between responses (which goes against the experimenter demand effect) than eliciting decisions on different screens.

MTurk sample. A concerned reader might be worried that our experimental setup is too complicated for an online implementation. We argue that this is unlikely to drive the data for several reasons. First, experiments involving beliefs elicitations and information processing have been successfully conducted on MTurk.[33] Nevertheless, to address this concern further, we replicated our analyses on a (pre-registered) sub-sample of participants from which we excluded participants who indicated in the post-experimental questionnaire that (i) the instructions were not extremely clear or that (ii) they had trouble understanding the instructions. These analyses are reported in the Appendix Sect. C.2 and show that our main findings are robust—and sometimes stronger—in the restricted sample.

7 Conclusion

Other-regarding preferences are prevalent in most human societies. However, the robustness of these preferences tends to be challenged in the presence of uncertainty in the decision environment. For instance, individuals with outcome-based preferences have been shown to exploit uncertainty about the relationship between their actions and outcomes to behave more selfishly. In contrast, the literature on belief-dependent preferences has focused on situations where the uncertainty about others' expectations is automatically resolved when the action is implemented. Hence, one can wonder whether individuals with belief-based preferences strategically acquire information about other's expectations to minimize the tension between their monetary incentives and their other-regarding preferences.

[33] For instance, Exley and Kessler (2021) replicated the original study by Dana et al. (2007) on information avoidance on MTurk and found results that are highly consistent with the original lab study.

To address this question, we adapted the information acquisition model proposed by Spiekermann and Weiss (2016) to study whether agents with belief-dependent preferences strategically acquire information about others' expectations. Our model predicts that agents with objective belief-dependent preferences always prefer more information, while agents with subjective belief-dependent preferences strategically seek information that minimizes the tension between their monetary interest and their other-regarding preferences. We then tested our predictions in an online experiment. We designed a modified trust game in which we manipulate trustees' beliefs about trustors' expectations by varying trustors' outside options. We then elicited trustees' preferences by asking them to report their return choices conditionally on the trustors' outside option. Finally, trustees were given the opportunity to acquire information about the trustors' outside option.

We found that 60.47% of trustees classified as belief-dependent engaged in self-serving information acquisition by choosing to acquire only the signal that was congruent with their monetary incentives, which is consistent with our theoretical predictions for subjective preferences. These findings suggest that previous research may have captured an upper bound of the positive impact of belief-dependent preferences on pro-social behavior. Interestingly, we found no evidence that the individuals with the most to gain from engaging in self-serving information acquisition (i.e., the ones with the most differentiated return decisions) were the most likely to do so. Finally, it is worth mentioning that a non-trivial fraction of our sample acquired information in a pattern consistent with objective belief-dependent preferences (20.93%).

Our findings underline the challenge of designing effective information policies to promote pro-sociality. We show that nudging belief-dependent individuals toward pro-social choices requires that information on others' expectations is attended to. When information is available but not directly observable, individuals may be tempted to seek self-serving signals, leading to less pro-social behavior than one would expect when information on expectations cannot be ignored.

Appendix A Proofs of the Theory

A.1 Predictions for a belief-discordant trustee

Following the same reasoning as in Sect. 2, we investigate the information acquisition strategy of belief-dependent trustees with belief-discordant preferences under uncertainty. Let $\hat{u}_i(\phi_\omega) = \max_y u_i(y, \phi_\omega)$ be the maximum utility achievable for a belief-discordant trustee for a given expectation ϕ_ω. In contrast with belief-concordant trustees, this function increases with ϕ: the higher the expectations of the trustor, the more likely it is that the trustee will return $y = 0$. The proof is provided in the Appendix A.2. Recalling our auxiliary assumption which states that $\phi_L < \phi_H$, it follows that $\hat{u}_i(\phi_L) < \hat{u}_i(\phi_H)$.

As in Sect. 2.2, when the trustor's expectation is uncertain, we distinguish between objective and subjective belief-discordant trustees. An *objective* belief-discordant trustee must maximize the pleasure from a mismatch given by:

$p \cdot c(y_U, \phi_L) + (1-p) \cdot c(y_U, \phi_H)$. Hence, the information acquisition strategy that maximizes his utility is to acquire both signals, as it maximizes his chances to learn about the trustor's actual expectation. The proof is provided in the Appendix A.4. In contrast, a subjective belief-discordant trustee maximizes his utility when he believes that the trustor's expectation is High (recall that, $\hat{u}_i(\phi_H) > \hat{u}_i(\phi_L)$). Consequently, a *subjective* belief-discordant trustee will only sample information from the signal S_H only, which provides either information congruent with this belief, or no information. The proof is provided in Appendix A.5. We summarize these results in the two following propositions.

Proposition 4 *An objective belief-discordant trustee acquires both signals.*

Proposition 5 *A subjective belief-discordant trustee acquires a High signal only.*

To summarize, belief-discordant trustees with objective preferences will exhibit the same information acquisition strategy than belief-concordant trustees with objective preferences. In contrast, belief-discordant trustees with subjective preferences will acquire information from the High signal only while belief-concordant trustees with subjective preferences will acquire information from the Low signal only.

A.2 Proof on how to determine trustees' optimal return

We first show that the problem has only 3 solutions: $y = 0$, $y = \phi_\omega$ and $y = E$.

Recall that $u_i(y, \phi_\omega) = E - y - \gamma(|\phi_\omega - y|)$. Hence,

$$\frac{du_i(y,\phi_\omega)}{dy} = -1 - \gamma_i\left(-\left[\frac{\phi_\omega - y}{|\phi_\omega - y|}\right]\right)$$
$$\Leftrightarrow \frac{du_i(y,\phi_\omega)}{dy} = \gamma_i\left[\frac{\phi_\omega - y}{|\phi - y|}\right] - 1$$

We distinguish between 3 cases:

$$\text{If } y = \phi_\omega, \ \frac{du_i(y,\phi_\omega)}{dy} = 0$$
$$\text{If } y < \phi_\omega, \ \frac{du_i(y,\phi_\omega)}{dy} = \gamma_i - 1$$
$$\text{If } y > \phi_\omega, \ \frac{du_i(y,\phi_\omega)}{dy} = -\gamma_i - 1$$

when $\gamma_i \neq \{-1, 1\}$, it is straightforward that $u_i(y, \phi_\omega)$ is maximised for $y = \phi_\omega$. If $y \neq \phi_\omega$, the two corners solutions are $y^* = 0$ and $y^* = E$. When $\gamma_i = 1$ or $\gamma_i = -1$, then the solution is undetermined: $y^* \in [0, E]$.

We can now use comparative statics to determine how y^* depends on γ_i and ϕ_ω.

- $u_i(y=0,\phi_\omega) = E - \gamma_i\phi_\omega$
- $u_i(y=\phi_\omega,\phi_\omega) = E - \phi_\omega$
- $u_i(y=E,\phi_\omega) = -\gamma_i(|\phi_\omega - E|) \Leftrightarrow -\gamma_i(E-\phi_\omega)$ since $\phi_\omega \leq E$

We first compare $u_i(y,\phi_\omega)$ when $y=0$ and when $y=\phi_\omega$:

$$\begin{aligned} u_i(y=0,\phi_\omega) >& u_i(y=\phi_\omega,\phi_\omega)\\ \Leftrightarrow& E-\gamma_i\phi_\omega > E-\phi_\omega\\ \Leftrightarrow& \gamma_i < 1 \end{aligned}$$

We then compare $u_i(y,\phi_\omega)$ when $y=\phi_\omega$ and when $y=E$:

$$\begin{aligned} u_i(y=\phi_\omega,\phi_\omega) >& u_i(y=E,\phi_\omega)\\ \Leftrightarrow& E-\phi_\omega > -\gamma(E-\phi_\omega)\\ \Leftrightarrow& \gamma_i > -1 \end{aligned}$$

We can see that (i) when $\gamma_i > 1$, a trustee prefers to return $y=\phi_\omega$ instead of $y=0$ or $y=E$, and (ii) when $-1 < \gamma_i < 1$, a trustee prefers to return $y=0$ instead of $y=\phi_\omega$, and hence instead of $y=E$.

It remains unclear what will the trustee prefer when $\gamma_i < -1$. To derive predictions in this case, we compare $u_i(y,\phi_\omega)$ when $y=0$ and when $y=E$:

$$\begin{aligned} u_i(y=0,\phi_\omega) >& u_i(y=E,\phi_\omega)\\ \Leftrightarrow& E-\gamma_i\phi_\omega > -\gamma_i(E-\phi_\omega)\\ \Leftrightarrow& E-\gamma_i\phi_\omega > -\gamma_iE+\gamma_i\phi_\omega\\ \Leftrightarrow& E-\gamma_i\phi_\omega+\gamma_iE-\gamma_i\phi_\omega > 0\\ \Leftrightarrow& \gamma(E-2\phi_\omega) > -E\\ \Leftrightarrow& \begin{cases} \gamma_i & > & -\frac{E}{E-2\phi_\omega} \text{ if } \phi_\omega > \frac{E}{2}\\ \gamma_i & < & -\frac{E}{E-2\phi_\omega} \text{ if } \phi_\omega < \frac{E}{2} \end{cases} \end{aligned}$$

when $\phi_\omega = \frac{E}{2}$, the solution is undetermined and $y \in [0,E]$.

When $\phi_\omega > \frac{E}{2}$, $-\frac{E}{E-2\phi_\omega} > 0$. Hence, the inequality $\gamma_i < -\frac{E}{E-2\phi_\omega}$ never holds. Hence, the trustee always prefer to return $y=0$ instead of $y=E$.

When $\phi_\omega < \frac{E}{2}$, the inequality $\gamma_i > -\frac{E}{E-2\phi_\omega}$ holds for some combinations of γ_i and ϕ_ω. More specifically, as ϕ_ω increases, $-\frac{E}{E-2\phi_\omega}$ also decreases in ϕ_ω over $[0,\frac{E}{2}[$. This means that, the higher ϕ_ω, the more likely it is that γ_i is above the threshold, the more likely it is that $u_i(y=0,\phi_\omega) > u_i(y=E,\phi_\omega)$, and the more likely it is that the trustee will return $y=0$ instead of $y=E$.

To summarise, we can conclude that

- When $\gamma_i = 1$ or $\gamma_i = -1$, the solution is undetermined: $y^* \in [0,E]$.
- When $\gamma_i > 1$, $y^* = \phi_\omega$
- When $-1 < \gamma_i < 1$, $y^* = 0$

- When $\gamma_i < -1$:
  - If $\phi_\omega > \frac{E}{2}$, $y^* = 0$
  - If $\phi_\omega < \frac{E}{2}$, $y^* = 0$ or $y^* = E$ but as ϕ_ω increases, the trustee becomes more likely to return $y^* = 0$ instead of $y^* = E$
  - If $\phi_\omega = \frac{E}{2}$, the solution is undetermined: $y^* \in [0, E]$

A.3 Proof on the variation of $\hat{u}$ with respect to ϕ_ω

According to the envelope theorem, the total derivative at point y^* is equal to the following partial derivative:

$$\begin{aligned}\hat{u}_i'(\phi_\omega) =&\frac{\partial}{\partial\phi_\omega}u_i(y,\phi_\omega)\Big|_{y=y^*}\\ =&\frac{\partial}{\partial\phi_\omega}[(E-y^*)-c(y^*,\phi_\omega)]\\ =&\frac{\partial}{\partial\phi_\omega}[(E-y^*)-\gamma_i\cdot(|\phi_\omega-y^*|)]\\ =&-\gamma_i\end{aligned}$$

Hence, we can conclude that $\hat{u}'_{i|\gamma>1}(\phi_\omega) \leq 0$ and $\hat{u}'_{i|\gamma<-1}(\phi_\omega) \geq 0$.

A.4 Proof of Proposition 2

Let $v(y)$ represent $E - y$. Let y^*_ω, with $\omega \in \{L, U, H\}$, represent the amount that maximizes a belief-concordant trustee's (expected) utility.

When acquiring the signal is S_L, the expected utility of a belief-concordant trustee with objective preferences is the weighted sum of the trustee's utility when the true state is $\omega = L$ and the trustee knows it (with probability ps), when the true state is S_L but the trustee does not know it (with probability $p(1-s)$), and when the true state is S_H (with probability $(1-p)$).

$$\begin{aligned}Eu_L =&ps\cdot u_{i,|\gamma>1}(y^*_L,\phi_L)+p(1-s)\cdot v(y^*_U,\phi_L)-p(1-s)\cdot c(y^*_U,\phi_L)\\ &+(1-p)\cdot v(y^*_U,\phi_H)-(1-p)\cdot c(y^*_U,\phi_H)\\ =&ps\cdot u_{i,|\gamma>1}(y^*_L,\phi_L)+(1-ps)\cdot v(y^*_U)-p(1-s)\cdot c(y^*_U,\phi_L)\\ &-(1-p)\cdot c(y^*_U,\phi_H)\end{aligned} \tag{5}$$

Similarly, when acquiring the signal is S_L, the expected utility of an objective belief-discordant trustee is given by the following equation.

$$\begin{aligned}Eu_L =&ps\cdot u_{i,|\gamma<-1}(y^*_L,\phi_L)+(1-ps)\cdot v(y^*_U)+p(1-s)\cdot c(y^*_U,\phi_L)\\ &+(1-p)\cdot c(y^*_U,\phi_H)\end{aligned} \tag{6}$$

When acquiring the signal is S_H, the expected utility of a trustee with objective belief-dependent preferences is the weighted sum of the trustee's utility when the true state is $\omega = H$ and the trustee knows it (with probability $(1-p)s$), when the true state is S_H but the trustee does not know it (with probability $(1-p)(1-s)$), and when the true state is S_L (with probability p).

$$\begin{aligned} Eu_H =&(1-p)s \cdot u_{i,|\gamma>1}(y_H^*, \phi_H) + (1-p)(1-s) \cdot v(y_U^*, \phi_H) \\ &- (1-p)(1-s) \cdot c(y_U^*, \phi_H) + p \cdot v(y_U^*, \phi_L) - p \cdot c(y_U^*, \phi_L) \\ =&(1-p)s \cdot u_{i,|\gamma>1}(y_H^*, \phi_H) + (1-s+ps) \cdot v(y_U^*) \\ &- (1-p)(1-s) \cdot c(y_U^*, \phi_H) \\ &- p \cdot c(y_U^*, \phi_L) \end{aligned} \tag{7}$$

Similarly, when acquiring the signal is S_H, the expected utility of an objective belief-discordant trustee is given by the following equation.

$$\begin{aligned} Eu_H =&(1-p)s \cdot u_{i,|\gamma<-1}(y_H^*, \phi_H) + (1-s+ps) \cdot v(y_U^*) \\ &+ (1-p)(1-s) \cdot c(y_U^*, \phi_H) \\ &+ p \cdot c(y_U^*, \phi_L) \end{aligned} \tag{8}$$

When acquiring both signals, the expected utility of a belief-concordant trustee with objective belief-dependent preferences is the weighted sum of the trustee's utility when the true state is $\omega = L$ and the trustee knows it (with probability ps), when the true state is $\omega = H$ and the trustee knows it (with probability $(1-p)s$) when the true state is S_L but the trustee does not know it (with probability $p(1-s)$), and when the true state is S_H but the trustee does not know it (with probability $((1-p)(1-s))$.

$$\begin{aligned} Eu_{LH} =&ps \cdot u_{i,|\gamma>1}(y_L^*, \phi_L) + (1-p)s \cdot u_{i,|\gamma>1}(y_H^*, \phi_H) \\ &+ p(1-s) \cdot v(y_U^*, \phi_L) - p(1-s) \cdot c(y_U^*, \phi_L) \\ &+ (1-p)(1-s) \cdot v(y_U^*, \phi_H) - (1-p)(1-s) \cdot c(y_U^*, \phi_H) \\ =&ps \cdot u_{i,|\gamma>1}(y_L^*, \phi_L) + (1-p)s \cdot u_{i,|\gamma>1}(y_H^*, \phi_H) + (1-s) \cdot v(y_U^*) \\ &- p(1-s) \cdot c(y_U^*, \phi_L) - (1-p)(1-s) \cdot c(y_U^*, \phi_H) \end{aligned} \tag{9}$$

Similarly, when acquiring both signals, the expected utility of an objective belief-discordant trustee is given by the following equation.

$$\begin{aligned} Eu_{LH} =&ps \cdot u_{i,|\gamma<-1}(y_L^*, \phi_L) + (1-p)s \cdot u_{i,|\gamma<-1}(y_H^*, \phi_H) \\ &+ (1-s) \cdot v(y_U^*) + p(1-s) \cdot c(y_U^*, \phi_L) \\ &+ (1-p)(1-s) \cdot c(y_U^*, \phi_H) \end{aligned} \tag{10}$$

We compare the expected utilities of receiving signal S_H (Eu_H) to receiving both signals (Eu_{LH}) for a belief-concordant trustee.

$$\begin{aligned} Eu_{LH} - Eu_H =& ps \cdot u_{i,|\gamma>1}(y_L^*, \phi_L) + (1-p)s \cdot u_{i,|\gamma>1}(y_H^*, \phi_H) \\ &+ (1-s) \cdot v(y_U^*) - p(1-s) \cdot c(y_U^*, \phi_L) \\ &- (1-p)(1-s) \cdot c(y_U^*, \phi_H) - (1-p)s \cdot u_{i,|\gamma>1}(y_H^*, \phi_H) \\ &- (1-s+ps) \cdot v(y_U^*) \\ &+ (1-p)(1-s) \cdot c(y_U^*, \phi_H) + p \cdot c(y_U^*, \phi_L) \\ =& ps \cdot [u_{i,|\gamma>1}(y_L^*, \phi_L) - v(y_U^*) + c(y_U^*, \phi_L)] \\ =& ps \cdot \underbrace{[u_{i,|\gamma>1}(y_L^*, \phi_L) - u_i(y_U^*, \phi_L)]}_{>0} \end{aligned} \tag{11}$$

Eq. 11 is positive since, given ϕ_L, utility is maximal at $\hat{u}_{i,|\gamma>1}(\phi_L) = u_{i,|\gamma>1}(y_L^*, \phi_L)$. Using the same reasoning, it yields to the following equation for subjective belief-discordant trustees.

$$Eu_{LH} - Eu_H = ps \cdot \underbrace{[u_{i,|\gamma>-1}(y_L^*, \phi_L) - u_{i,|\gamma>-1}(y_U^*, \phi_L)]}_{>0} \tag{12}$$

We compare the expected utilities of receiving signal S_L (Eu_L) to receiving both signals (Eu_{LH}) for a belief-concordant trustee.

$$\begin{aligned} Eu_{LH} - Eu_L =& ps \cdot u_{i,|\gamma>1}(y_L^*, \phi_L) + (1-p)s \cdot u_{i,|\gamma>1}(y_H^*, \phi_H) \\ &+ (1-s) \cdot v(y_U^*) - p(1-s) \cdot a_i(y_U^*, \phi_L) \\ &- (1-p)(1-s) \cdot g_i(y_U^*, \phi_H) - ps \cdot u_{i,|\gamma>1}(y_L^*, \phi_L) \\ &- (1-ps) \cdot v(y_U^*) \\ &+ p(1-s) \cdot c(y_U^*, \phi_L) + (1-p) \cdot c(y_U^*, \phi_H) \\ =& (1-p)s \cdot [u_{i,|\gamma>1}(y_H^*, \phi_H) - u_{i,|\gamma>1}(y_U^*, \phi_H)] > 0 \end{aligned} \tag{13}$$

Eq. 13 is positive since, given ϕ_H, utility is maximal at $\hat{u}_{i,|\gamma>1}(\phi_H) = u_{i,|\gamma>1}(y_H^*, \phi_H)$. Using the same reasoning, it yields to the following equation for subjective expectation-based reciprocal trustees.

$$Eu_{LH} - Eu_L = (1-p)s \cdot [u_{i,|\gamma<-1}(y_H^*, \phi_H) - u_{i,|\gamma<-1}(y_U^*, \phi_H)] > 0 \tag{14}$$

We can conclude that taking both signals is the preferred choice for both objective belief-concordant and objective belief-discordant trustees.

A.5 Proof of Proposition 3

The expected utility of acquiring signal S_L for a for a belief-concordant trustee with subjective preferences corresponds to the weighted sum of the trustee's utility when the state is $\omega = L$ and the trustee knows it (with probability ps), and when the trustee is uncertain about the state (with probability $1 - ps$).

$$Eu_L = ps \cdot \hat{u}_{i,|\gamma>1}(\Phi_L) + (1 - ps) \cdot \hat{u}_{i,|\gamma>1}(\Phi_U) \tag{15}$$

Symmetrically, The expected utility of acquiring signal S_H for a subjective belief-concordant trustee corresponds to the weighted sum of the trustee's utility when the state is $\omega = H$ and the trustee knows it (with probability $(1 - p)s$), and when the trustee is uncertain about the state (with probability $1 - s + ps$).

$$Eu_H = (1 - p)s \cdot \hat{u}_i(\Phi_H) + (1 - s + ps) \cdot \hat{u}_i(\Phi_U) \tag{16}$$

Finally, the expected utility of acquiring both signal for a subjective belief-concordant trustee corresponds to the weighted sum of the trustee's utility when the state is $\omega = L$ and the trustee knows it (with probability ps), when the state is $\omega = H$ and the trustee knows it (with probability $(1 - p)s$), and when the trustee is uncertain about the state (with probability $1 - s$).

$$Eu_{LH} = ps \cdot \hat{u}_i(\Phi_L) + (1 - p)s \cdot \hat{u}_i(\Phi_H) + (1 - s) \cdot \hat{u}_i(\Phi_U) \tag{17}$$

First, we focus on the case of subjective belief-concordant trustees. To conclude from the equations below, recall that (i) since $\Phi_L < \Phi_U < \Phi_H$, it follows that $\hat{u}_{i,|\gamma>1}(\Phi_L) > \hat{u}_{i,|\gamma>1}(\Phi_U) > \hat{u}_{i,|\gamma>1}(\Phi_H)$, and (ii) p and $s \in [0, 1]$.

$$\begin{aligned} Eu_L - Eu_H =& ps \cdot \hat{u}_{i,|\gamma>1}(\Phi_L) + (1 - ps) \cdot \hat{u}_{i,|\gamma>1}(\Phi_U) \\ & - (1 - p)s \cdot \hat{u}_{i,|\gamma>1}(\Phi_H) - (1 - s + ps) \cdot \hat{u}_{i,|\gamma>1}(\Phi_U) \\ =& ps \cdot \hat{u}_{i,|\gamma>1}(\Phi_L) + (1 - ps - 1 + s - ps) \\ & \cdot \hat{u}_{i,|\gamma>1}(\Phi_U) - (1 - p)s \cdot \hat{u}_{i,|\gamma>1}(\Phi_H) \\ =& ps \cdot \underbrace{[\hat{u}_{i,|\gamma>1}(\Phi_L) - \hat{u}_{i,|\gamma>1}(\Phi_U)]}_{>0} + (1 - p)s \\ & \cdot \underbrace{[\hat{u}_{i,|\gamma>1}(\Phi_U) - \hat{u}_{i,|\gamma>1}(\Phi_H)]}_{>0} \end{aligned} \tag{18}$$

$$\begin{aligned} Eu_L - Eu_{LH} =& ps \cdot \hat{u}_{i,|\gamma>1}(\Phi_L) + (1 - ps) \cdot \hat{u}_{i,|\gamma>1}(\Phi_U) \\ & - ps \cdot \hat{u}_{i,|\gamma>1}(\Phi_L) - (1 - p)s \cdot \hat{u}_{i,|\gamma>1}(\Phi_H) - (1 - s) \cdot \hat{u}_{i,|\gamma>1}(\Phi_U) \\ =& (ps - ps) \cdot \hat{u}_{i,|\gamma>1}(\Phi_L) + (1 - ps - 1 + s) \\ & \cdot \hat{u}_{i,|\gamma>1}(\Phi_U) - (1 - p)s \cdot \hat{u}_{i,|\gamma>1}(\Phi_H) \\ =& s(1 - p) \cdot \underbrace{[\hat{u}_{i,|\gamma>1}(\Phi_U) - \hat{u}_{i,|\gamma>1}(\Phi_H)]}_{>0} > 0 \end{aligned} \tag{19}$$

We can conclude that, under uncertainty, a subjective belief-concordant trustee who follows a coarse mapping, will acquire signal S_L, but neither signal S_H nor both signals.

We now turn to the case of subjective belief-discordant trustees. To conclude from the equations below, recall that (i) since $\Phi_L < \Phi_U < \Phi_H$, it follows that $\hat{u}_{i,|\gamma<-1}(\Phi_L) < \hat{u}_{i,|\gamma<-1}(\Phi_U) < \hat{u}_{i,|\gamma<-1}(\Phi_H)$, and (ii) p and $s \in [0, 1]$.

$$\begin{aligned} Eu_H - Eu_L =& (1-p)s \cdot \hat{u}_{i,|\gamma<-1}(\Phi_H) + (1-s+ps) \cdot \hat{u}_{i,|\gamma<-1}(\Phi_U) \\ & - ps \cdot \hat{u}_{i,|\gamma<-1}(\Phi_L) - (1-ps) \cdot \hat{u}_{i,|\gamma<-1}(\Phi_U) \\ =& (s-ps) \cdot \hat{u}_{i,|\gamma<-1}(\Phi_H) + (1-s+ps-1+ps) \\ & \cdot \hat{u}_{i,|\gamma<-1}(\Phi_U) - ps \cdot \hat{u}_{i,|\gamma<-1}(\Phi_L) \\ =& (s-ps) \cdot \underbrace{[\hat{u}_{i,|\gamma<-1}(\Phi_H) - \hat{u}_{i,|\gamma<-1}(\Phi_U)]}_{>0} + ps \\ & \cdot \underbrace{[\hat{u}_{i,|\gamma<-1}(\Phi_U) - \hat{u}_{i,|\gamma<-1}(\Phi_L)]}_{>0} \end{aligned} \tag{20}$$

$$\begin{aligned} Eu_H - Eu_{LH} =& (1-p)s \cdot \hat{u}_{i,|\gamma<-1}(\Phi_H) + (1-s+ps) \\ & \cdot \hat{u}_{i,|\gamma<-1}(\Phi_U) - ps \cdot \hat{u}_{i,|\gamma<-1}(\Phi_L) \\ & - (1-p)s \cdot \hat{u}_{i,|\gamma<-1}(\Phi_H) - (1-s) \cdot \hat{u}_{i,|\gamma<-1}(\Phi_U) \\ =& (1-s+ps-1+s) \cdot \hat{u}_{i,|\gamma<-1}(\Phi_U) - ps \cdot \hat{u}_{i,|\gamma<-1}(\Phi_L) \\ =& ps \cdot \underbrace{[\hat{u}_{i,|\gamma<-1}(\Phi_U) - \hat{u}_{i,|\gamma<-1}(\Phi_L)]}_{>0} \end{aligned} \tag{21}$$

We can conclude that, under uncertainty, a subjective belief-discordant trustee who follows a coarse mapping, will acquire signal S_H, but neither signal S_L nor both signals.

Appendix B Additional results

B.1 Summary statistics

See Table 2.

Table 2 Summary statistics

	Mean	Standard errors	N
Female	0.40	.	320
Age	40.21	0.683	320
Clarity	1.52	0.036	320
Request in Beauty Contest	54.70	1.073	320

Clarity takes value “1” when instructions were deemed “extremely clear” and “4” when the instructions were “extremely unclear”

B.2 Aggregated beliefs

Figure 6 shows that the trustors' median belief is lower in the Low game (median = 60 cents; interquartile range = 75)[34] than in the High game (med = 90 cents; iqr = 15). This difference is significant at the 0.01% level (Wilcoxon rank-sum test, $p < 0.001$).[35] Similarly, the trustees' median belief about trustors' belief is lower in the Low game (med = 75 cents; iqr = 60) than in the High game (med = 90 cents; iqr = 30). This difference is also significant at the 0.01% level (Wilcoxon signed rank test, $p < 0.001$).

B.3 Trustor's behavior

We showed in Sect. 5.1 that trustor's expect to receive more from the trustees when their outside option is high rather than low. Consistent with Eq. 1, 91.82% of trustors who choose to go *In* expects to receive at least their outside option. Moreover, the share of trustors choosing *In* is lower when the outside option is High (51.85%) rather than Low (86.08%) (chi-square test, $p < 0.001$).

B.4 Trustees' justification of their sampling strategies

We classified the participants' justification of their sampling strategies in four categories (excluding 11 trustees who did not fill in this optional question, and those who did not satisfy our auxiliary assumption on beliefs). The first category pools the trustees who made their choice out of curiosity, e.g., "*I was just curious to see if I would find a 15 or 75*". Second, we grouped together participants who mentioned their intention to maximize their payoff, e.g., "*I chose to open 1 silver envelope hoping it would contain a 15 and then I would maximize my earnings*". In the third category, we pooled the participants who reported having made their choice at random, e.g., "*I chose 1 envelope honestly just based on feeling*". The last category contains answers that we could not classify in the other three categories.

Table 3a shows that when opening one envelope only, the majority of belief-dependent trustees choose at random; while they are motivated by curiosity when they open both envelopes. Table 3b shows that the majority opened a silver envelope to maximize their payoff, while they opened both envelope to satisfy their curiosity.

B.5 Trustees' likelihood of having a given sampling strategy

Table 4 reports the average marginal effect of a multinomial logit model using a categorical variable equals to 0 if the trustee opened a silver envelope only, 1 if the trustee opened a gold envelop only and 2 if the trustee opened both a silver and a gold

[34] Respectively med and iqr, hereafter.

[35] All p-values are two-sided.

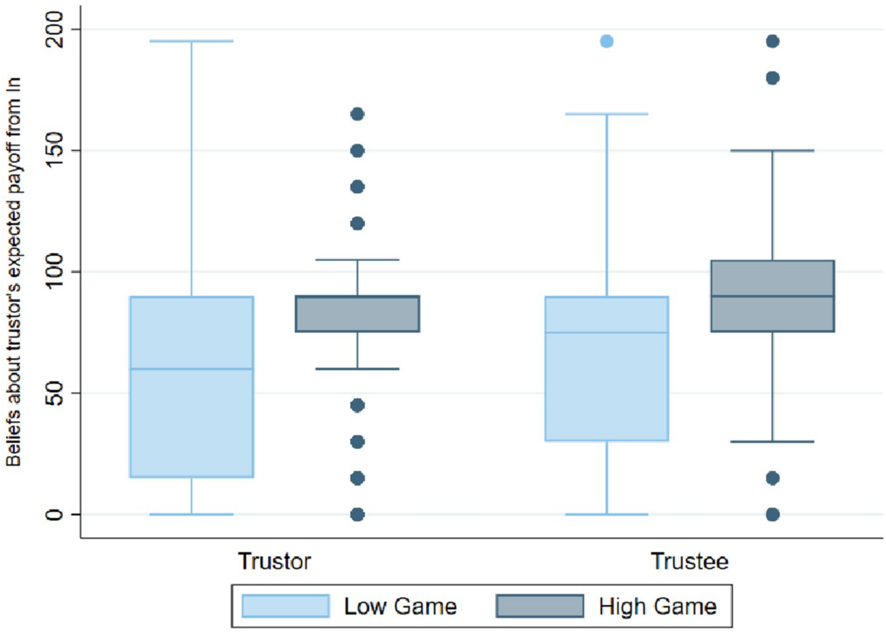


Fig. 6 Distribution of trustors and trustees' beliefs about trustors' payoff from *In*

envelopes as the dependent variable. Regressors include a dummy variable equal to 1 if the trustee is belief-concordant, and 0 if the trustee is belief-independent. belief-concordant trustees are more likely to open a silver envelope and less likely to open both envelopes than belief-independent trustees and the results are significant at the 0.1% level. These results are robust to the inclusion of individual controls.

To investigate the driver of differences in trustee's information acquisition strategy, we replicate Table 1 from the main text after including belief-independent subjects. The average marginal effects (AME) are displayed in Table 5. We found that an increase in 10 cents in the difference in conditional returns increases the likelihood to open a silver envelope by up to 9 percentage points (columns (1) and (2)), and decreases the likelihood to open a gold envelope by up to 9 percentage points (columns (5) and (6)). These findings show that individuals who have the most money to lose from learning about a specific state of the world, are also the ones who are the most likely to engage in self-serving information acquisition strategies.

B.6 Determinants of beliefs, returns and preference type

To investigate the determinants of participants' beliefs, we estimated a linear regression of the difference in beliefs for both trustors and trustees on participants' individual characteristics. The OLS coefficients are displayed in columns (1) and (2) in Table 6, respectively. We find that an increase in the perceived clarity of the

Table 3 Trustees' justification of their sampling strategies

	Curiosity (%)	Payoff (%)	Random (%)	Other (%)	Total (n)
(a) Belief-independent trustees					
Open Silver	12.50	25.00	50.00	12.50	8
Open Gold	36.36	9.09	36.36	18.18	11
Open Both	62.50	4.17	20.83	12.50	24
(b) belief-concordant trustees					
Open Silver	4.17	79.17	4.17	12.50	24
Open Gold	12.50	37.50	12.50	37.50	8
Open Both	75.00	12.50	0.00	12.50	8

Table 4 Average marginal effects of preferences types on the likelihood of each sampling strategy

Belief-independent	Open Silver		Open Gold		Open both	
	Ref	Ref	Ref	Ref	Ref	Ref
belief-concordant	0.428***	0.453***	– 0.108	– 0.105	– 0.320***	– 0.348***
	(0.092)	(0.086)	(0.087)	(0.085)	(0.093)	(0.085)
Ind. controls	No	Yes	No	Yes	No	Yes
Observations	94	94	94	94	94	94

This Table reports the average marginal effects estimated by Multinomial Logit models. Individual controls include the amount guessed in the beauty contest game, and socio-demographic characteristics (age, gender, annual income, weekly expenditure). Standard errors are in parentheses

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Table 5 Average marginal effects of monetary incentives on the likelihood of each sampling strategy

	Open Silver		Open Gold		Open both	
	(1)	(2)	(3)	(4)	(5)	(6)
Return(High)–Return(Low)	0.008***	0.009***	−0.001	−0.000	−0.007**	−0.009***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Ind. controls	No	Yes	No	Yes	No	Yes
Observations	94	94	94	94	94	94

Table 5 reports the average marginal effects of our multinomial logit model of the difference in conditional returns on the likelihood of a given sampling strategy. Controls include the amount guessed in the beauty contest game and socio-demographic characteristics (age, identifying as a female, annual income, weekly expenditure). Standard errors are in parentheses

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

instructions increases trustor's sensitivity to our treatment manipulation ($p = 0.038$), but not trustees'. Surprisingly, we find no effect of participants' guess in the beauty contest on their sensitivity to the treatment manipulation ($p = 0.557$ and $p = 0.536$).

In addition, we investigate the determinants of trustees' difference in conditional return choices. To do so, we estimated a linear regression of the difference in conditional returns on trustees' individual characteristics. The OLS coefficients

Table 6 Determinants of participants' beliefs, trustees' conditional return decisions and preferences type

Dep. var	Diff. belief	Diff. belief	Diff. return	Types Trustees	
	Trustors	Trustees	Trustees	Belief Ind	belief-concordant
	(1)	(2)	(3)	(4)	(5)
Level of reasoning	0.078	−0.077	−0.141	0.001	−0.002
	(0.133)	(0.125)	(0.135)	(0.003)	(0.003)
Female	−2.502	5.807	1.464	−0.051	0.019
	(4.933)	(5.198)	(5.169)	(0.112)	(0.111)
Age	−0.226	−0.321	−0.106	0.008	−0.005
	(0.209)	(0.212)	(0.238)	(0.005)	(0.005)
Annual income	0.306	−1.551	0.226	0.0108	−0.003
	(1.975)	(1.957)	(1.986)	(0.043)	(0.042)
Weekly expenditure	−4.128	6.009	−0.315	0.002	−0.016
	(3.629)	(3.277)	(3.263)	(0.070)	(0.069)
Clarity instructions	8.619*	2.721	−2.137	0.122	−0.110
	(4.114)	(3.621)	(3.771)	(0.083)	(0.081)
Constant	42.19***	31.67*	21.79	–	–
	(13.38)	(14.17)	(14.34)		
Observations	160	160	98	98	98

Table 6 displays the OLS coefficients of participants' individual characteristics on trustors' (column (1)) and trustees' (column (2)) differences in beliefs between the Low and the High game, trustees' differences in return between the Low and the High game (column (3)), as well as the marginal effect from a logit regression of trustees' individual characteristics on their preference type (columns (5) and (6)). Standard errors are in parentheses

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

are displayed in column (3). We find no effect of trustees' individual characteristics on their conditional return choices.

Finally, we investigate the determinants of trustees' preference type. To do so, we estimated logit regression of the likelihood of having belief-independent or belief-concordant preferences on trustees' individual characteristics. The average marginal effects are displayed in columns (4) and (5), respectively. We find no effect of trustees' individual characteristics on their preference type.

Appendix C Robustness checks

C.1 Replication analyses on the pooled sample (original and follow-up studies)

In this section, we replicate our main analyses on the pooled observations from both the original study and the follow-up study.

Are beliefs affected by the outside option manipulation?

We find that 57.73% of trustors and 59.62% of trustees hold higher beliefs in the High game than in the Low game. In contrast, 16.09% of trustors and 14.51% of trustees indicated higher beliefs in the Low game than in the high Game, while 26.18% of trustors and 25.87% of trustees indicated similar expectations regardless of the game being played. We conclude that Result 1 remains quantitatively the same.

Are trustees motivated by belief-dependent preferences?

We find that 38.52% (n=73) of trustees can be classified as belief-independent, 54.50% (n=103) of trustees can be classified as belief-concordant and 6.88% (n=13) of trustees can be classified as belief-discordant. Result 2 remains quantitatively the same.

How do belief-based preferences affect information acquisition?

We found that 44.66% of belief-concordant trustees chose to open a silver envelope only, 23.30% opened a gold envelope only, and 32.04% opened both envelopes. For beliefs-independent trustees, we found that 16.44% opened a silver envelope only, and 34.25% opened a gold envelope only, and 49.32% opened both envelopes. In addition, the proportion of belief-concordant trustees who chose to open a silver envelope only is significantly higher than the proportion of belief-independent trustees who made the same choice (Pearson's chi-square test, $p < 0.001$). Again, Result 3 is robust to the inclusion of the follow-up study.

C.2 Replication analyses on a restricted sample (original study)

We pre-registered that we will check the robustness of our findings by excluding from the analyses participants who indicated in the post-experimental questionnaire that (i) the instructions were not extremely clear or that (ii) the had trouble understanding the instructions. 59 trustors and 81 trustees indicated that the instructions were not extremely clear (43.75% of participants). In addition, 1 trustee indicated that they encountered comprehension problem with the instructions while indicating that the instructions were extremely clear (4.06% of participants). In the following section, we excluded these participants from the analyses.

Are beliefs affected by the outside option manipulation?

The proportion of participants who verifies Auxiliary Hypothesis 1 increases slightly. 61.39% (vs. 53.13%) of trustors and 64.10% (vs. 61.25%) of trustees hold higher beliefs in the High game than in the Low game. In contrast, 5.94% (vs. 10%) of trustors and 7.69% (vs. 8.13%) of trustees indicated higher beliefs in the Low game than in the high Game, while 32.67% (vs. 28.75%) of trustors and 28.21% (vs. 38.75%) of trustees indicated similar expectations regardless of the game being played. Overall, Result 1 is not affected by participants' comprehension of the experimental instructions.

Are trustees motivated by belief-dependent preferences?

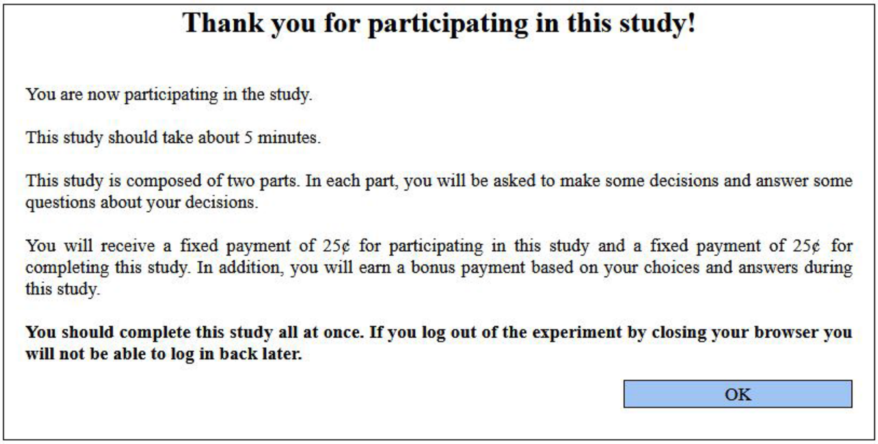
58% (n=29; vs. 52.04%, n=51) of trutees can be classified as belief-independent while 38% (n=19; vs. 43.88%, n = 43) of trustees can be classified as belief-concordant. Only two trustee can be classified as belief-discordant (vs. 4.08%, n = 4). Result 2 is not affected by participants' comprehension of the experimental instructions.

How do belief-based preferences affect information acquisition?

We found that 68.42% of belief-concordant trustees chose to open a silver envelope only, 10.53% opened a gold envelope only, and 21.05% opened both envelopes. For beliefs-independent trustees, we found that 17.24% opened a silver envelope only, and 31.03% opened a gold envelope only, and 51.72% opened both envelopes. In addition, the proportion of belief-concordant trustees who chose to open a silver envelope only is significantly higher than the proportion of belief-independent trustees who made the same choice (Pearson's chi-square test, $p < 0.001$). Again, Result 3 is not affected by participants' comprehension of the experimental instructions.

Appendix D Screens from the online experiment

D.1 Trustors' screens



Thank you for participating in this study!

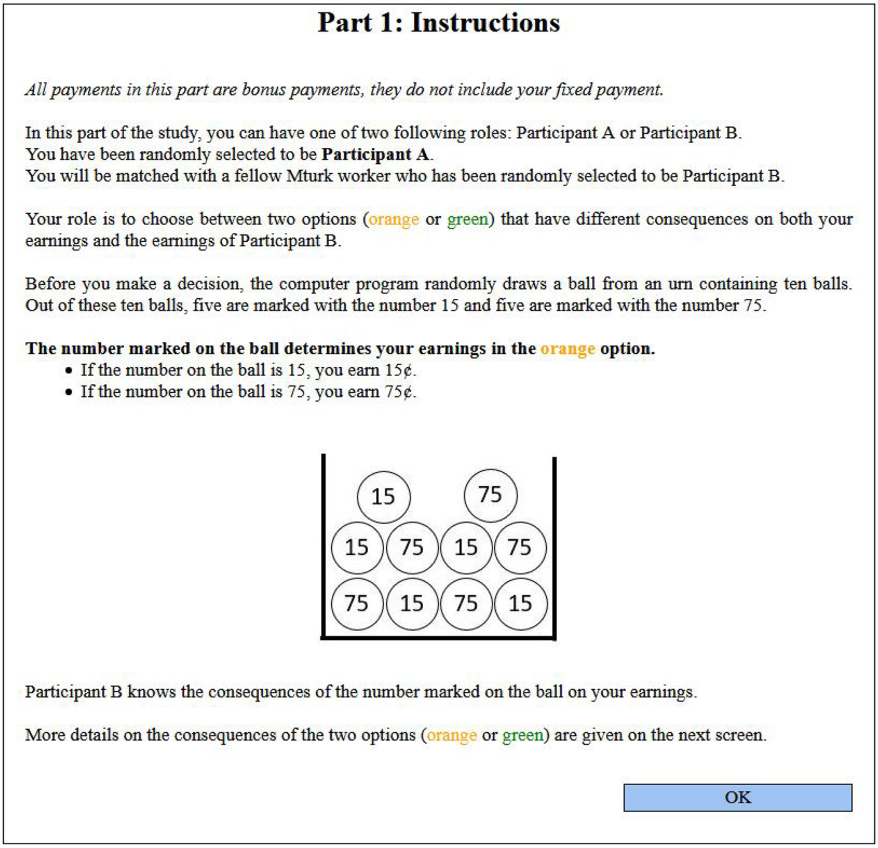
You are now participating in the study.

This study should take about 5 minutes.

This study is composed of two parts. In each part, you will be asked to make some decisions and answer some questions about your decisions.

You will receive a fixed payment of 25¢ for participating in this study and a fixed payment of 25¢ for completing this study. In addition, you will earn a bonus payment based on your choices and answers during this study.

You should complete this study all at once. If you log out of the experiment by closing your browser you will not be able to log in back later.



Part 1: Instructions

All payments in this part are bonus payments, they do not include your fixed payment.

In this part of the study, you can have one of two following roles: Participant A or Participant B.
You have been randomly selected to be **Participant A**.
You will be matched with a fellow Mturk worker who has been randomly selected to be Participant B.

Your role is to choose between two options (orange or green) that have different consequences on both your earnings and the earnings of Participant B.

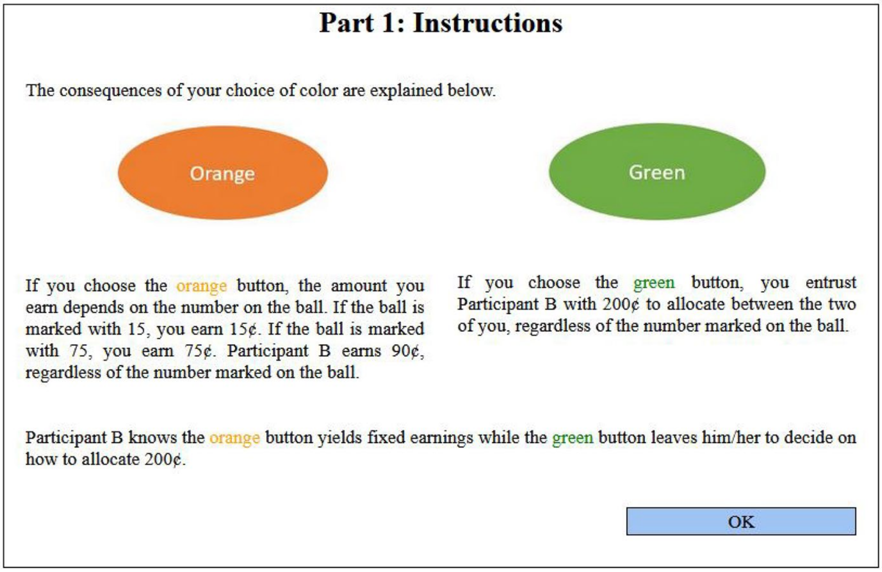
Before you make a decision, the computer program randomly draws a ball from an urn containing ten balls. Out of these ten balls, five are marked with the number 15 and five are marked with the number 75.

The number marked on the ball determines your earnings in the orange option.

- If the number on the ball is 15, you earn 15¢.
- If the number on the ball is 75, you earn 75¢.

Participant B knows the consequences of the number marked on the ball on your earnings.

More details on the consequences of the two options (orange or green) are given on the next screen.



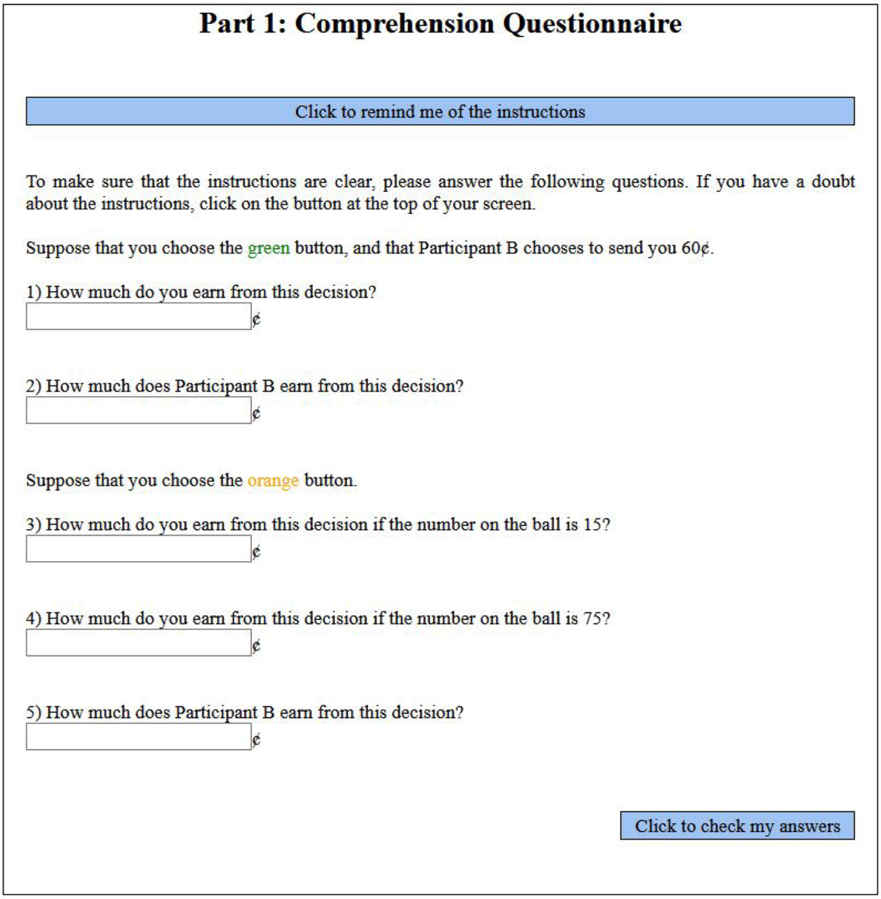
Part 1: Instructions

The consequences of your choice of color are explained below.

If you choose the orange button, the amount you earn depends on the number on the ball. If the ball is marked with 15, you earn 15¢. If the ball is marked with 75, you earn 75¢. Participant B earns 90¢, regardless of the number marked on the ball.

If you choose the green button, you entrust Participant B with 200¢ to allocate between the two of you, regardless of the number marked on the ball.

Participant B knows the orange button yields fixed earnings while the green button leaves him/her to decide on how to allocate 200¢.


Part 1: Comprehension Questionnaire

To make sure that the instructions are clear, please answer the following questions. If you have a doubt about the instructions, click on the button at the top of your screen.

Suppose that you choose the green button, and that Participant B chooses to send you 60¢.

1) How much do you earn from this decision?
______ ¢

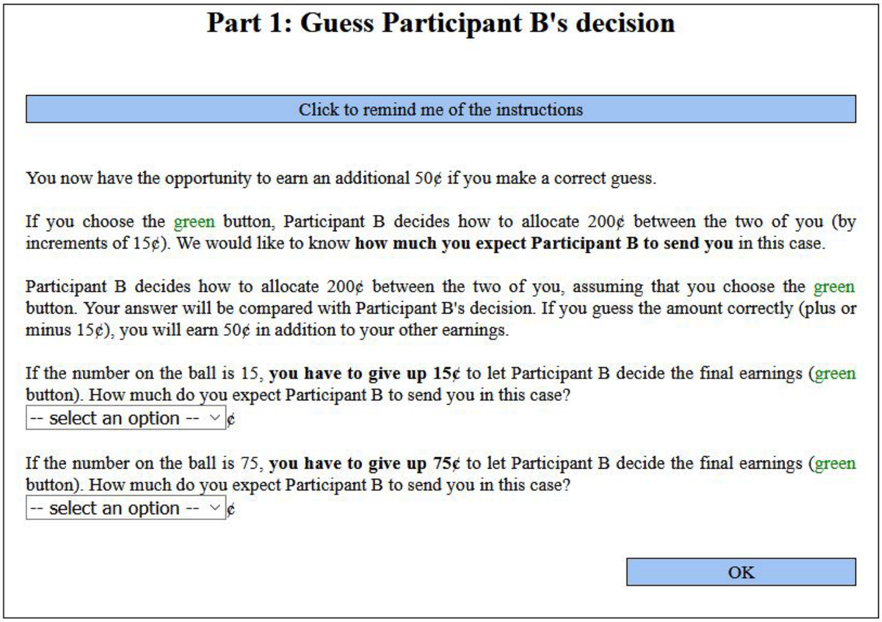
2) How much does Participant B earn from this decision?
______ ¢

Suppose that you choose the orange button.

3) How much do you earn from this decision if the number on the ball is 15?
______ ¢

4) How much do you earn from this decision if the number on the ball is 75?
______ ¢

5) How much does Participant B earn from this decision?
______ ¢



Part 1: Guess Participant B's decision

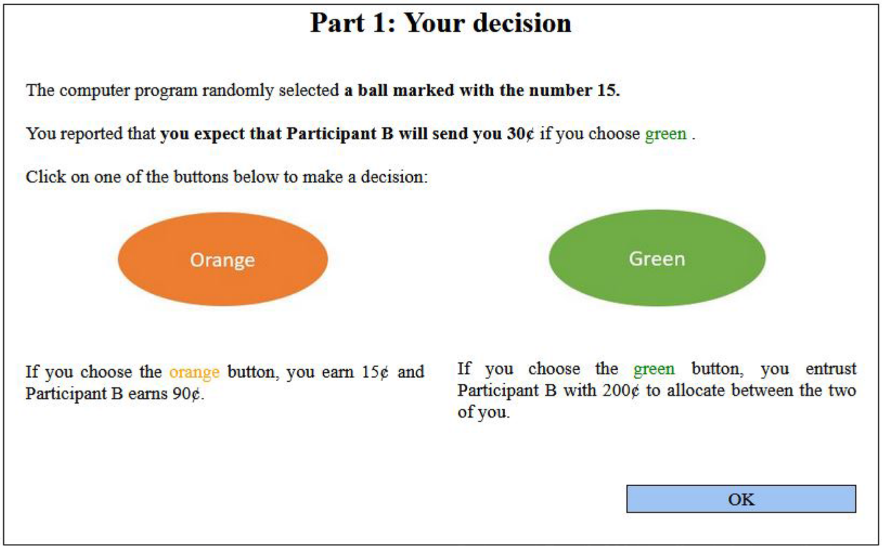
You now have the opportunity to earn an additional 50¢ if you make a correct guess.

If you choose the green button, Participant B decides how to allocate 200¢ between the two of you (by increments of 15¢). We would like to know **how much you expect Participant B to send you** in this case.

Participant B decides how to allocate 200¢ between the two of you, assuming that you choose the green button. Your answer will be compared with Participant B's decision. If you guess the amount correctly (plus or minus 15¢), you will earn 50¢ in addition to your other earnings.

If the number on the ball is 15, **you have to give up 15¢** to let Participant B decide the final earnings (green button). How much do you expect Participant B to send you in this case?
¢

If the number on the ball is 75, **you have to give up 75¢** to let Participant B decide the final earnings (green button). How much do you expect Participant B to send you in this case?
¢



Part 1: Your decision

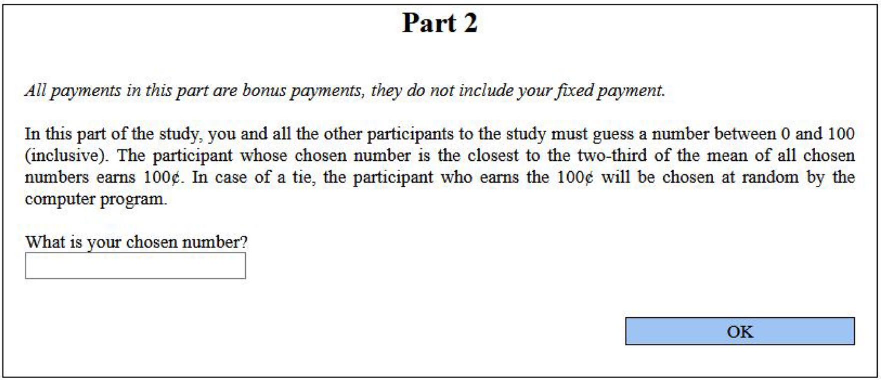
The computer program randomly selected **a ball marked with the number 15.**

You reported that **you expect that Participant B will send you 30¢** if you choose green .

Click on one of the buttons below to make a decision:

If you choose the orange button, you earn 15¢ and Participant B earns 90¢.

If you choose the green button, you entrust Participant B with 200¢ to allocate between the two of you.

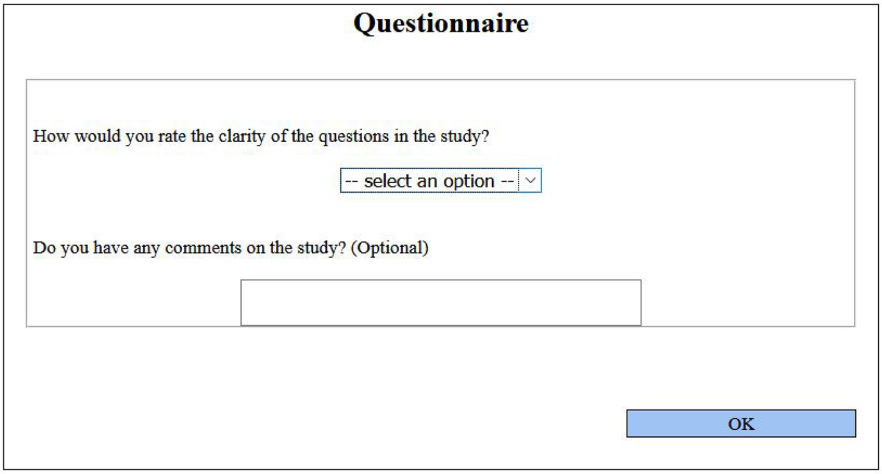


Part 2

All payments in this part are bonus payments, they do not include your fixed payment.

In this part of the study, you and all the other participants to the study must guess a number between 0 and 100 (inclusive). The participant whose chosen number is the closest to the two-third of the mean of all chosen numbers earns 100¢. In case of a tie, the participant who earns the 100¢ will be chosen at random by the computer program.

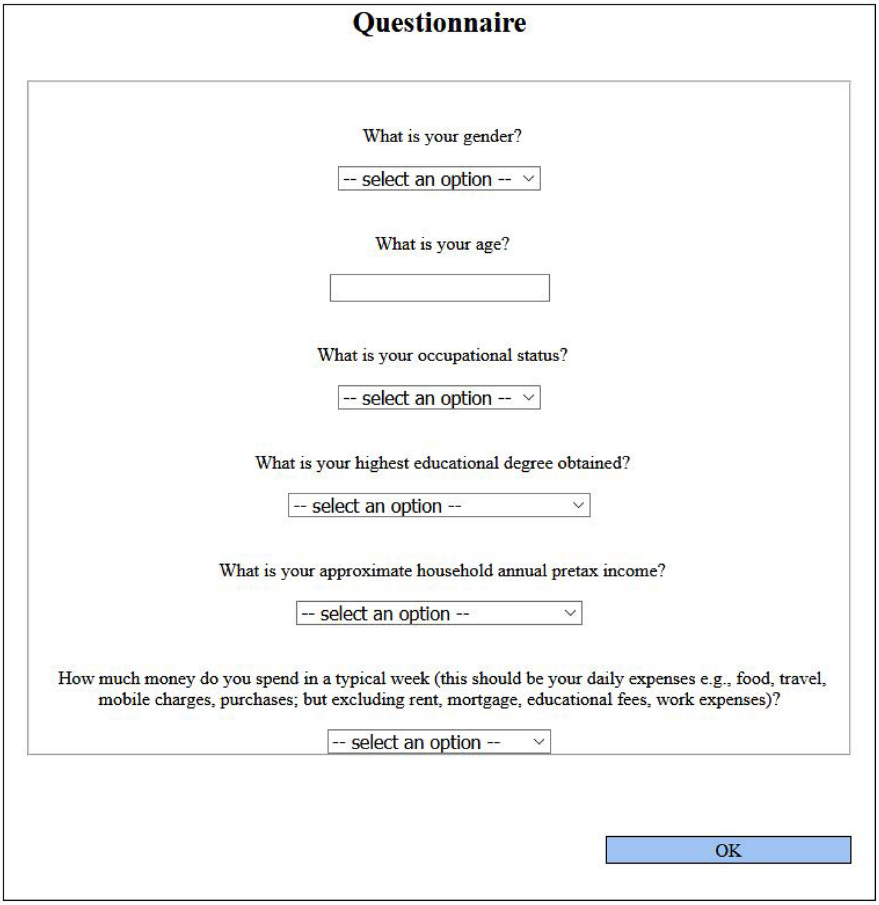
What is your chosen number?



Questionnaire

How would you rate the clarity of the questions in the study?

Do you have any comments on the study? (Optional)



Questionnaire

What is your gender?

What is your age?

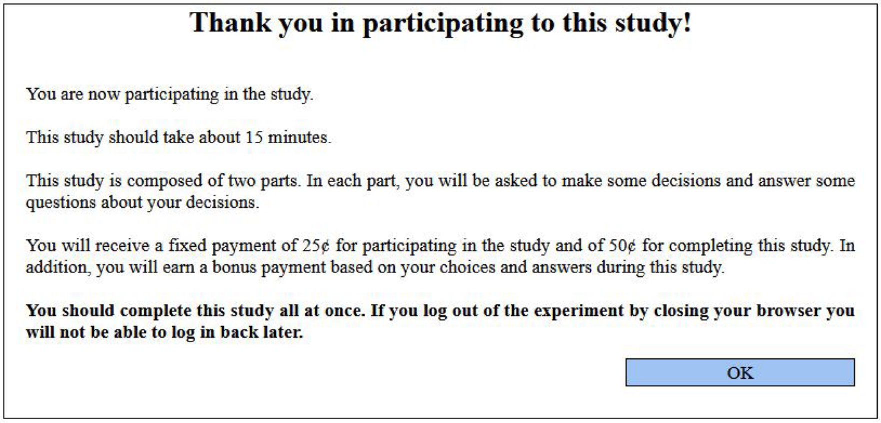
What is your occupational status?

What is your highest educational degree obtained?

What is your approximate household annual pretax income?

How much money do you spend in a typical week (this should be your daily expenses e.g., food, travel, mobile charges, purchases; but excluding rent, mortgage, educational fees, work expenses)?

D.2 Trustees' screens



Thank you in participating to this study!

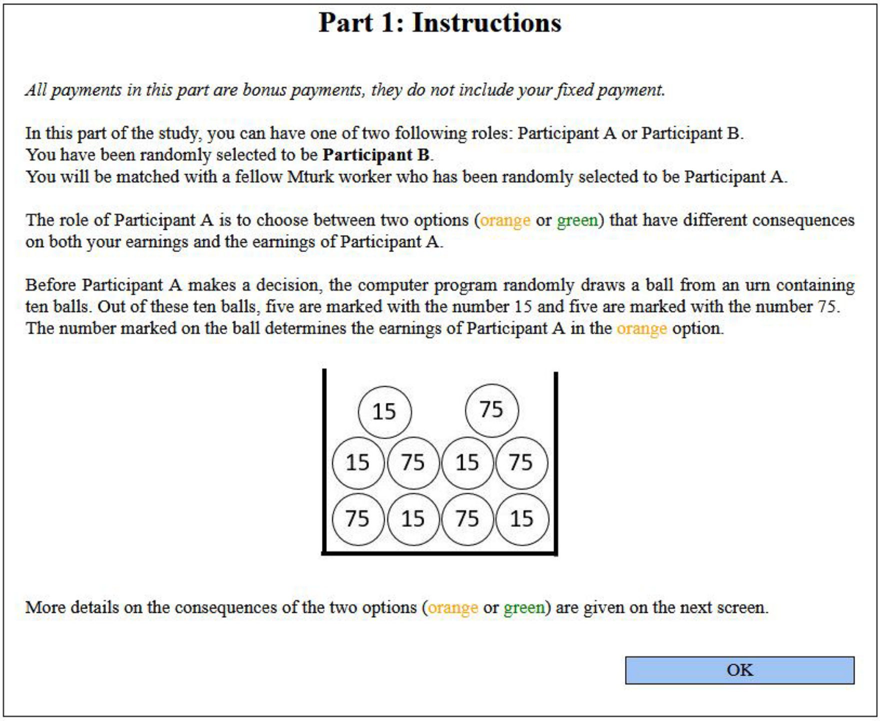
You are now participating in the study.

This study should take about 15 minutes.

This study is composed of two parts. In each part, you will be asked to make some decisions and answer some questions about your decisions.

You will receive a fixed payment of 25¢ for participating in the study and of 50¢ for completing this study. In addition, you will earn a bonus payment based on your choices and answers during this study.

You should complete this study all at once. If you log out of the experiment by closing your browser you will not be able to log in back later.



Part 1: Instructions

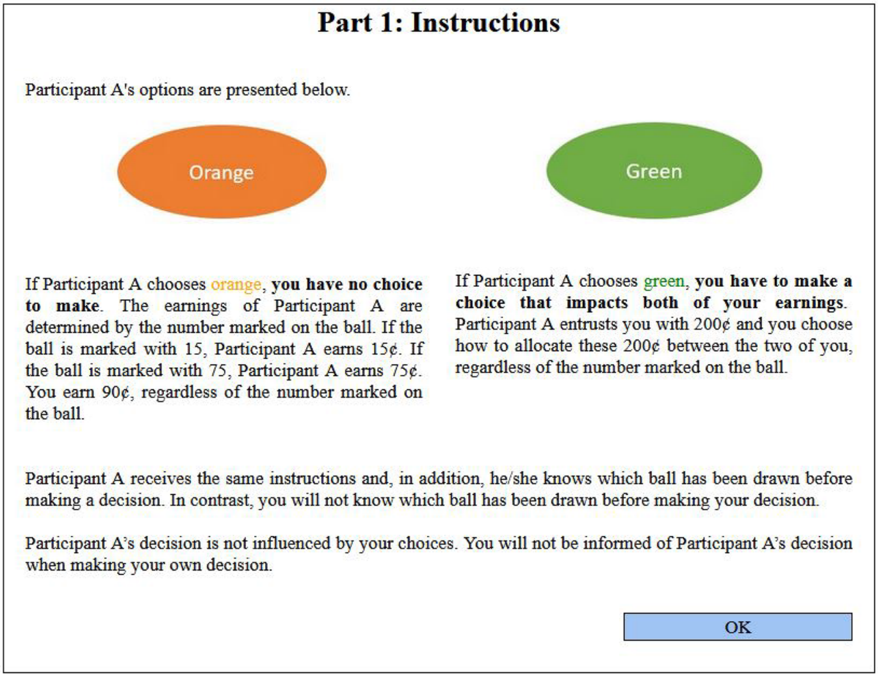
All payments in this part are bonus payments, they do not include your fixed payment.

In this part of the study, you can have one of two following roles: Participant A or Participant B.
You have been randomly selected to be **Participant B**.
You will be matched with a fellow Mturk worker who has been randomly selected to be Participant A.

The role of Participant A is to choose between two options (orange or green) that have different consequences on both your earnings and the earnings of Participant A.

Before Participant A makes a decision, the computer program randomly draws a ball from an urn containing ten balls. Out of these ten balls, five are marked with the number 15 and five are marked with the number 75. The number marked on the ball determines the earnings of Participant A in the orange option.

More details on the consequences of the two options (orange or green) are given on the next screen.



Part 1: Instructions

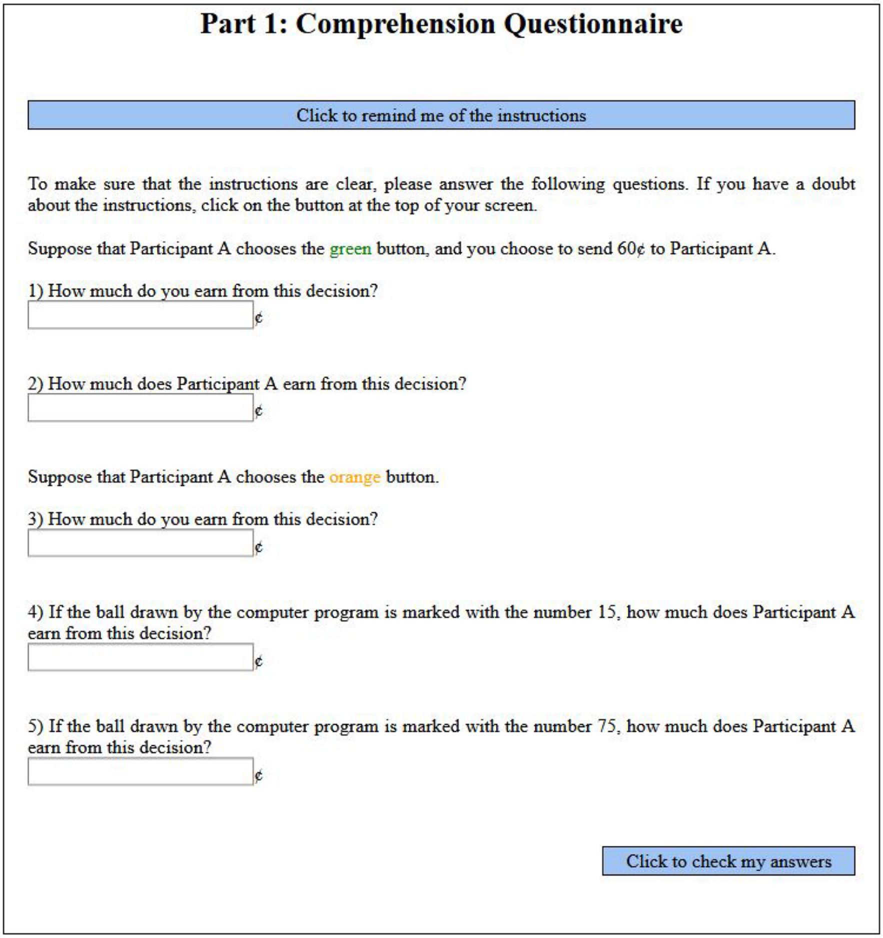
Participant A's options are presented below.

If Participant A chooses orange, **you have no choice to make**. The earnings of Participant A are determined by the number marked on the ball. If the ball is marked with 15, Participant A earns 15¢. If the ball is marked with 75, Participant A earns 75¢. You earn 90¢, regardless of the number marked on the ball.

If Participant A chooses green, **you have to make a choice that impacts both of your earnings**. Participant A entrusts you with 200¢ and you choose how to allocate these 200¢ between the two of you, regardless of the number marked on the ball.

Participant A receives the same instructions and, in addition, he/she knows which ball has been drawn before making a decision. In contrast, you will not know which ball has been drawn before making your decision.

Participant A's decision is not influenced by your choices. You will not be informed of Participant A's decision when making your own decision.


Click to remind me of the instructions
Click to check my answers

Part 1: Comprehension Questionnaire

To make sure that the instructions are clear, please answer the following questions. If you have a doubt about the instructions, click on the button at the top of your screen.

Suppose that Participant A chooses the green button, and you choose to send 60¢ to Participant A.

1) How much do you earn from this decision?
¢

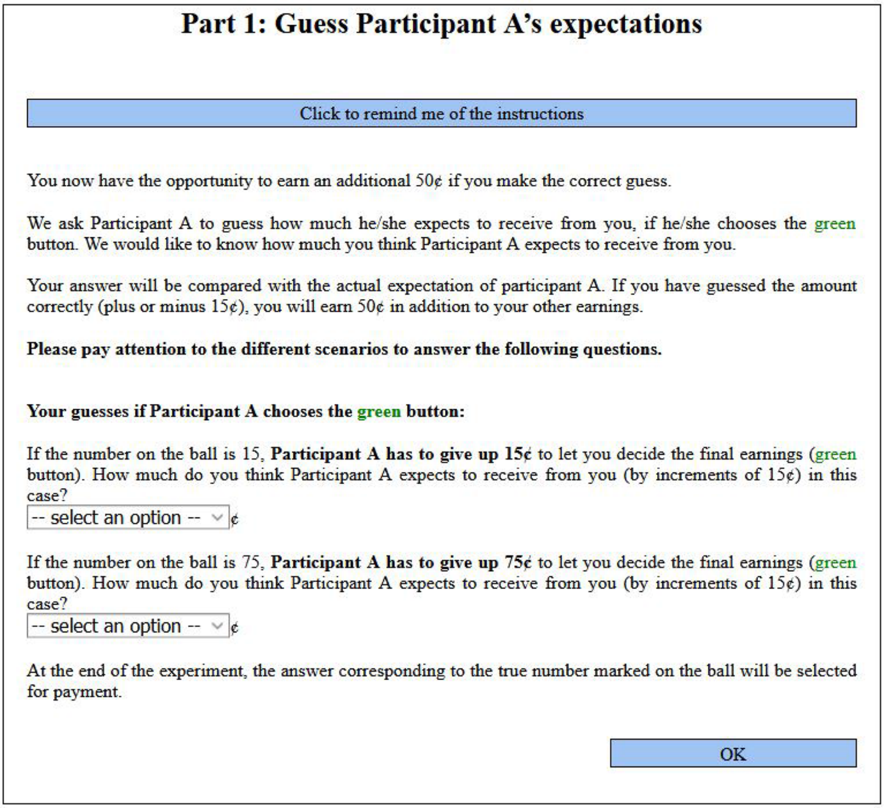
2) How much does Participant A earn from this decision?
¢

Suppose that Participant A chooses the orange button.

3) How much do you earn from this decision?
¢

4) If the ball drawn by the computer program is marked with the number 15, how much does Participant A earn from this decision?
¢

5) If the ball drawn by the computer program is marked with the number 75, how much does Participant A earn from this decision?
¢



Part 1: Guess Participant A's expectations

You now have the opportunity to earn an additional 50¢ if you make the correct guess.

We ask Participant A to guess how much he/she expects to receive from you, if he/she chooses the green button. We would like to know how much you think Participant A expects to receive from you.

Your answer will be compared with the actual expectation of participant A. If you have guessed the amount correctly (plus or minus 15¢), you will earn 50¢ in addition to your other earnings.

Please pay attention to the different scenarios to answer the following questions.

Your guesses if Participant A chooses the green button:

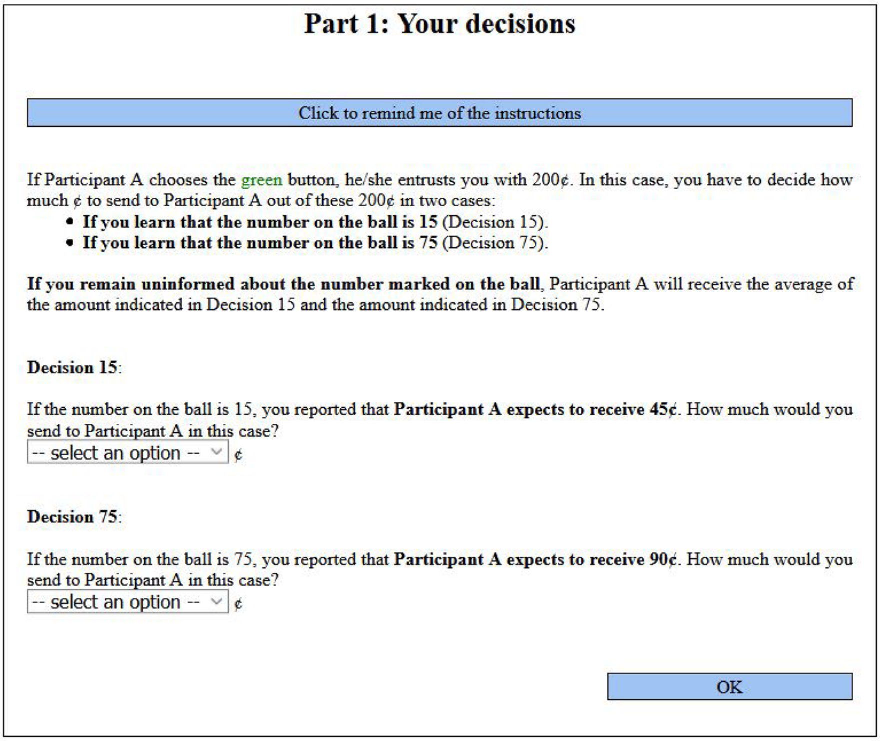
If the number on the ball is 15, **Participant A has to give up 15¢** to let you decide the final earnings (green button). How much do you think Participant A expects to receive from you (by increments of 15¢) in this case?

-- select an option -- ¢

If the number on the ball is 75, **Participant A has to give up 75¢** to let you decide the final earnings (green button). How much do you think Participant A expects to receive from you (by increments of 15¢) in this case?

-- select an option -- ¢

At the end of the experiment, the answer corresponding to the true number marked on the ball will be selected for payment.



Part 1: Your decisions

If Participant A chooses the green button, he/she entrusts you with 200¢. In this case, you have to decide how much ¢ to send to Participant A out of these 200¢ in two cases:

- **If you learn that the number on the ball is 15** (Decision 15).
- **If you learn that the number on the ball is 75** (Decision 75).

If you remain uninformed about the number marked on the ball, Participant A will receive the average of the amount indicated in Decision 15 and the amount indicated in Decision 75.

Decision 15:

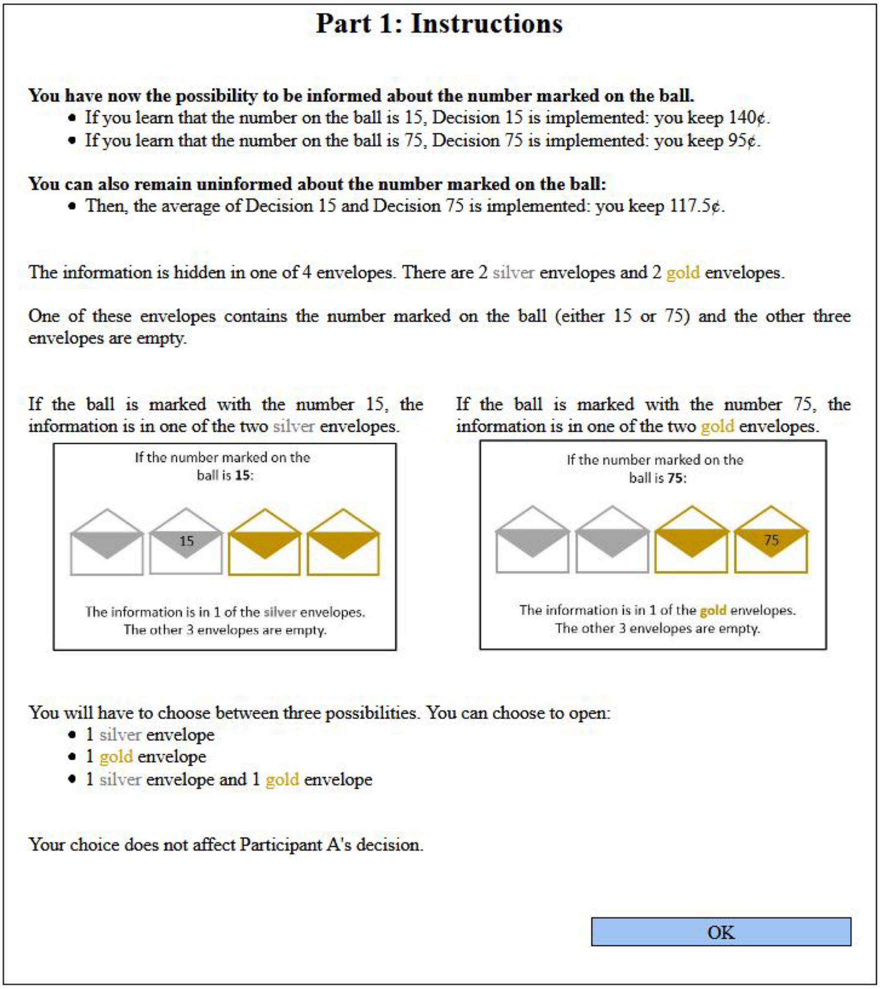
If the number on the ball is 15, you reported that **Participant A expects to receive 45¢**. How much would you send to Participant A in this case?

-- select an option -- ¢

Decision 75:

If the number on the ball is 75, you reported that **Participant A expects to receive 90¢**. How much would you send to Participant A in this case?

-- select an option -- ¢



Part 1: Instructions

You have now the possibility to be informed about the number marked on the ball.

- If you learn that the number on the ball is 15, Decision 15 is implemented: you keep 140¢.
- If you learn that the number on the ball is 75, Decision 75 is implemented: you keep 95¢.

You can also remain uninformed about the number marked on the ball:

- Then, the average of Decision 15 and Decision 75 is implemented: you keep 117.5¢.

The information is hidden in one of 4 envelopes. There are 2 silver envelopes and 2 gold envelopes.

One of these envelopes contains the number marked on the ball (either 15 or 75) and the other three envelopes are empty.

If the ball is marked with the number 15, the information is in one of the two silver envelopes.

If the number marked on the ball is **15**:

The information is in 1 of the silver envelopes. The other 3 envelopes are empty.

If the ball is marked with the number 75, the information is in one of the two gold envelopes.

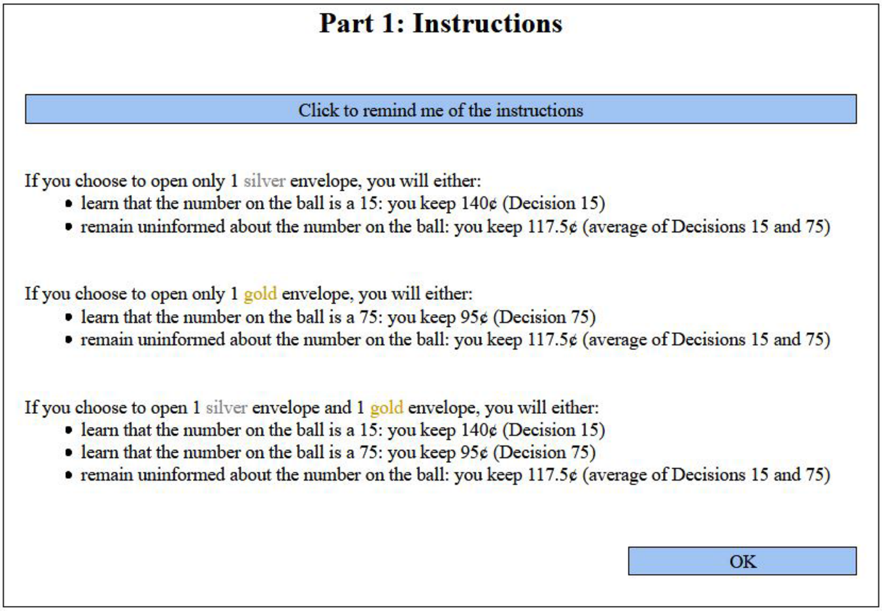
If the number marked on the ball is **75**:

The information is in 1 of the **gold** envelopes. The other 3 envelopes are empty.

You will have to choose between three possibilities. You can choose to open:

- 1 silver envelope
- 1 gold envelope
- 1 silver envelope and 1 gold envelope

Your choice does not affect Participant A's decision.



Part 1: Instructions

If you choose to open only 1 silver envelope, you will either:

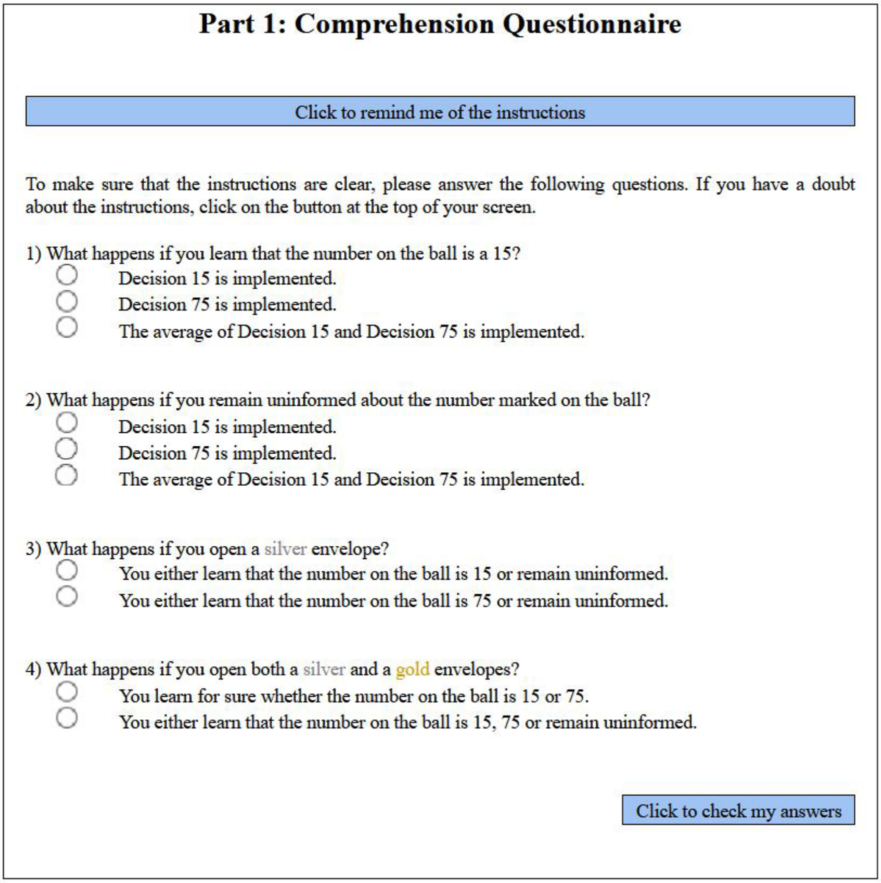
- learn that the number on the ball is a 15: you keep 140¢ (Decision 15)
- remain uninformed about the number on the ball: you keep 117.5¢ (average of Decisions 15 and 75)

If you choose to open only 1 gold envelope, you will either:

- learn that the number on the ball is a 75: you keep 95¢ (Decision 75)
- remain uninformed about the number on the ball: you keep 117.5¢ (average of Decisions 15 and 75)

If you choose to open 1 silver envelope and 1 gold envelope, you will either:

- learn that the number on the ball is a 15: you keep 140¢ (Decision 15)
- learn that the number on the ball is a 75: you keep 95¢ (Decision 75)
- remain uninformed about the number on the ball: you keep 117.5¢ (average of Decisions 15 and 75)



Part 1: Comprehension Questionnaire

To make sure that the instructions are clear, please answer the following questions. If you have a doubt about the instructions, click on the button at the top of your screen.

1) What happens if you learn that the number on the ball is a 15?

○ Decision 15 is implemented.
○ Decision 75 is implemented.
○ The average of Decision 15 and Decision 75 is implemented.

2) What happens if you remain uninformed about the number marked on the ball?

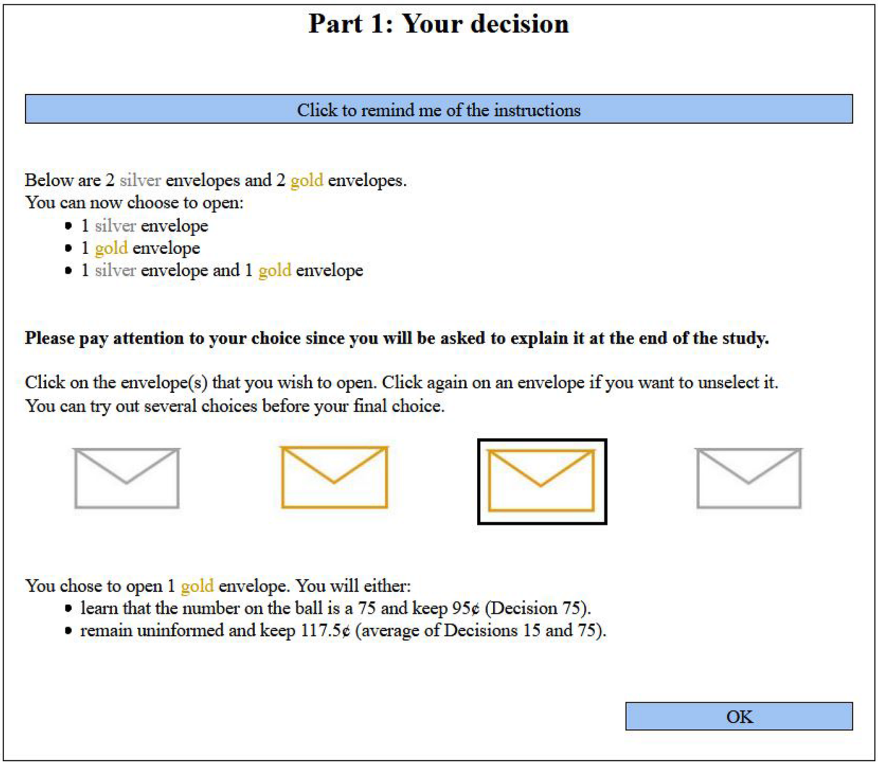
○ Decision 15 is implemented.
○ Decision 75 is implemented.
○ The average of Decision 15 and Decision 75 is implemented.

3) What happens if you open a silver envelope?

○ You either learn that the number on the ball is 15 or remain uninformed.
○ You either learn that the number on the ball is 75 or remain uninformed.

4) What happens if you open both a silver and a gold envelopes?

○ You learn for sure whether the number on the ball is 15 or 75.
○ You either learn that the number on the ball is 15, 75 or remain uninformed.



Part 1: Your decision

Below are 2 silver envelopes and 2 gold envelopes.
You can now choose to open:

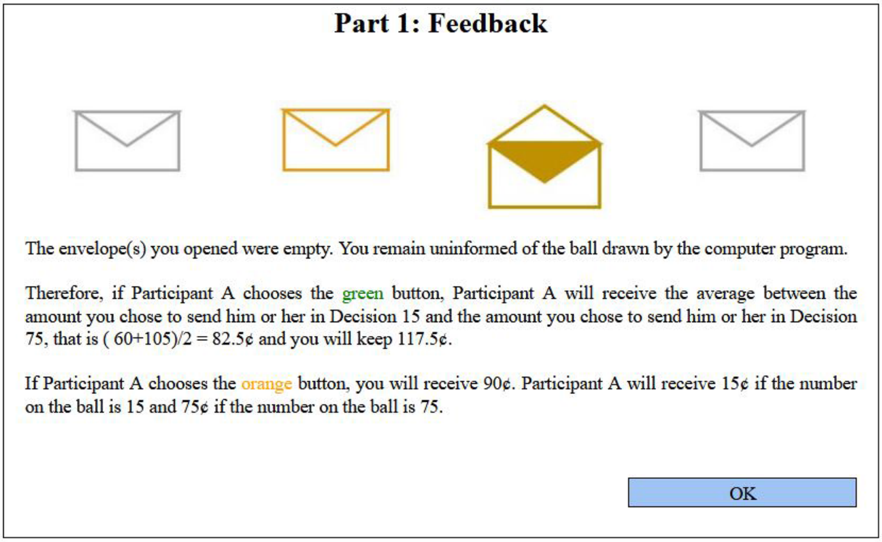
- 1 silver envelope
- 1 gold envelope
- 1 silver envelope and 1 gold envelope

Please pay attention to your choice since you will be asked to explain it at the end of the study.

Click on the envelope(s) that you wish to open. Click again on an envelope if you want to unselect it.
You can try out several choices before your final choice.

You chose to open 1 gold envelope. You will either:

- learn that the number on the ball is a 75 and keep 95¢ (Decision 75).
- remain uninformed and keep 117.5¢ (average of Decisions 15 and 75).

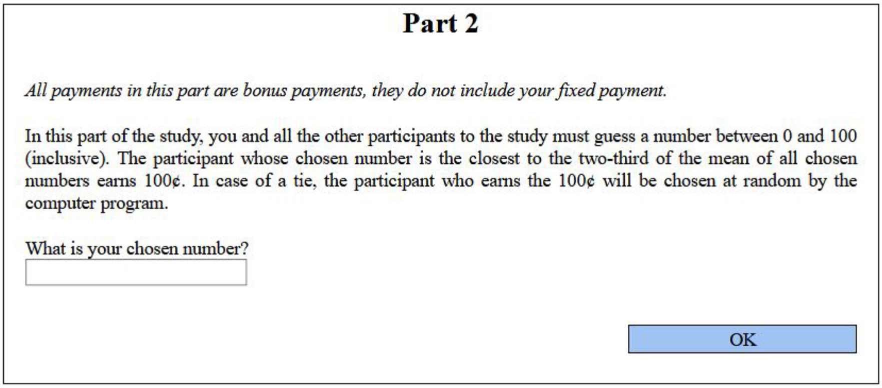


Part 1: Feedback

The envelope(s) you opened were empty. You remain uninformed of the ball drawn by the computer program.

Therefore, if Participant A chooses the green button, Participant A will receive the average between the amount you chose to send him or her in Decision 15 and the amount you chose to send him or her in Decision 75, that is (60+105)/2 = 82.5¢ and you will keep 117.5¢.

If Participant A chooses the orange button, you will receive 90¢. Participant A will receive 15¢ if the number on the ball is 15 and 75¢ if the number on the ball is 75.

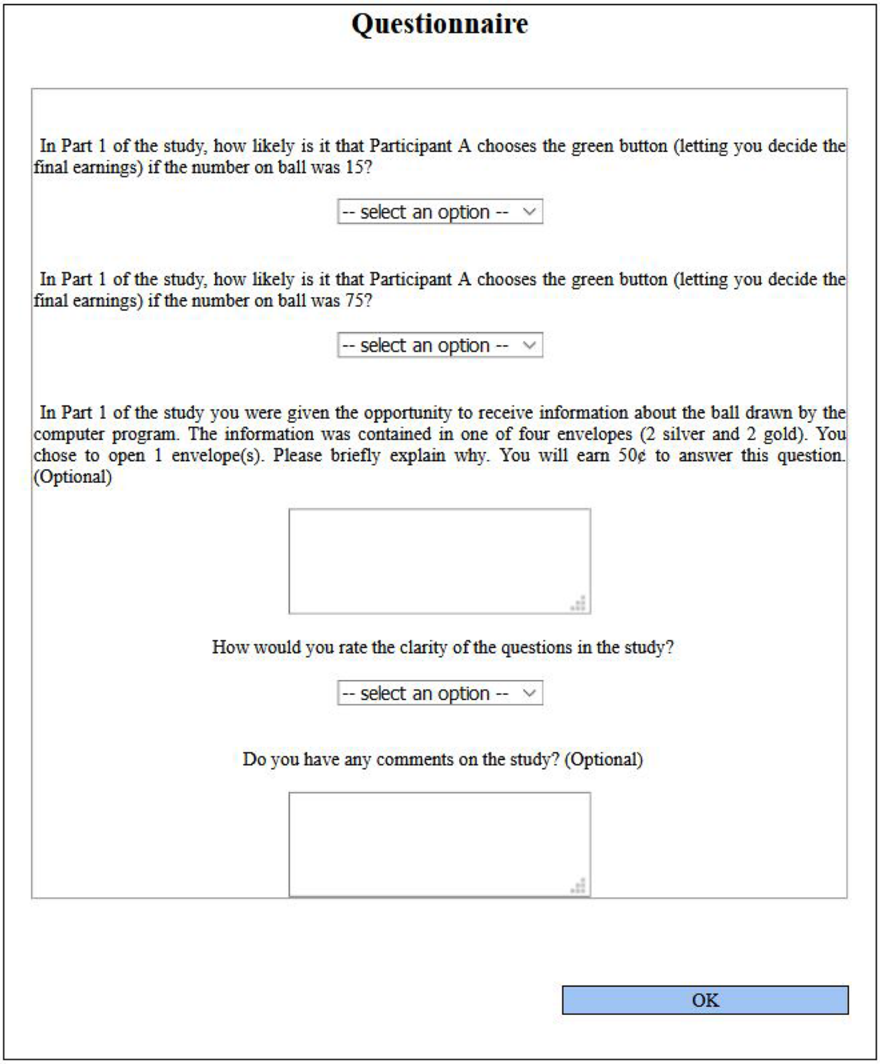


Part 2

All payments in this part are bonus payments, they do not include your fixed payment.

In this part of the study, you and all the other participants to the study must guess a number between 0 and 100 (inclusive). The participant whose chosen number is the closest to the two-third of the mean of all chosen numbers earns 100¢. In case of a tie, the participant who earns the 100¢ will be chosen at random by the computer program.

What is your chosen number?



Questionnaire

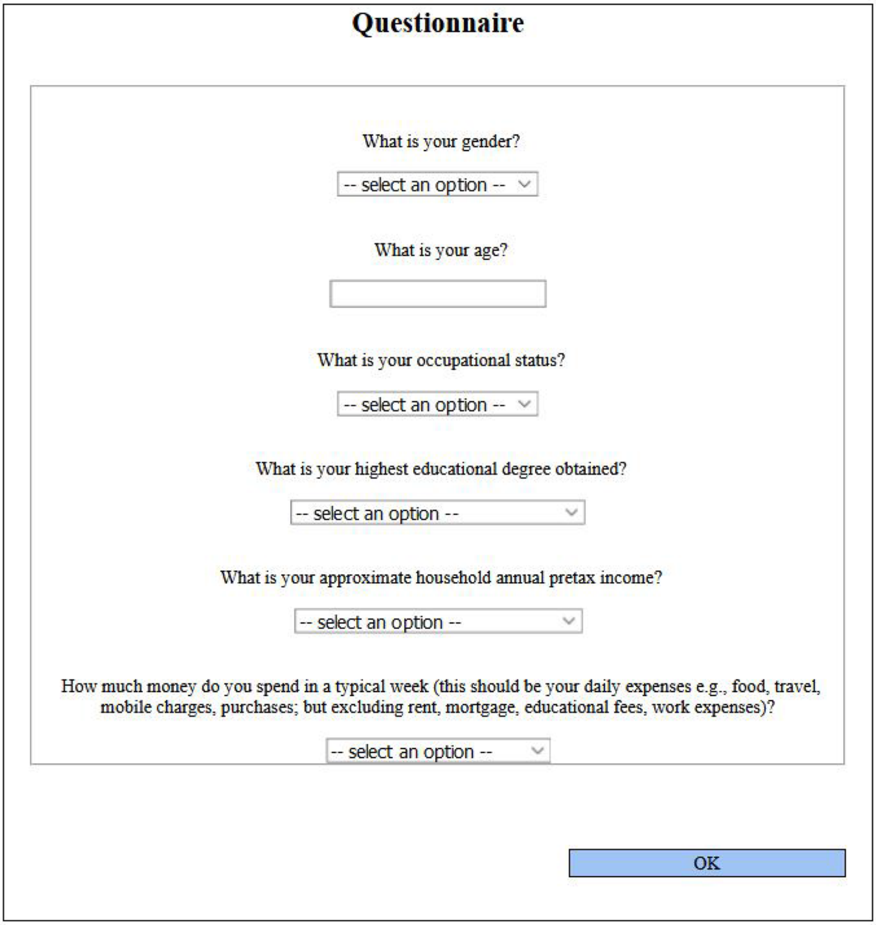
In Part 1 of the study, how likely is it that Participant A chooses the green button (letting you decide the final earnings) if the number on ball was 15?

In Part 1 of the study, how likely is it that Participant A chooses the green button (letting you decide the final earnings) if the number on ball was 75?

In Part 1 of the study you were given the opportunity to receive information about the ball drawn by the computer program. The information was contained in one of four envelopes (2 silver and 2 gold). You chose to open 1 envelope(s). Please briefly explain why. You will earn 50¢ to answer this question. (Optional)

How would you rate the clarity of the questions in the study?

Do you have any comments on the study? (Optional)



Questionnaire

What is your gender?

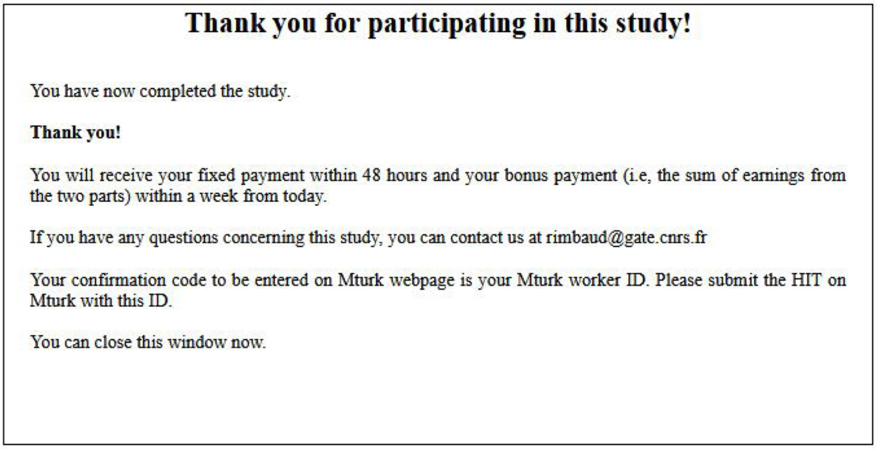
What is your age?

What is your occupational status?

What is your highest educational degree obtained?

What is your approximate household annual pretax income?

How much money do you spend in a typical week (this should be your daily expenses e.g., food, travel, mobile charges, purchases; but excluding rent, mortgage, educational fees, work expenses)?



Thank you for participating in this study!

You have now completed the study.

Thank you!

You will receive your fixed payment within 48 hours and your bonus payment (i.e, the sum of earnings from the two parts) within a week from today.

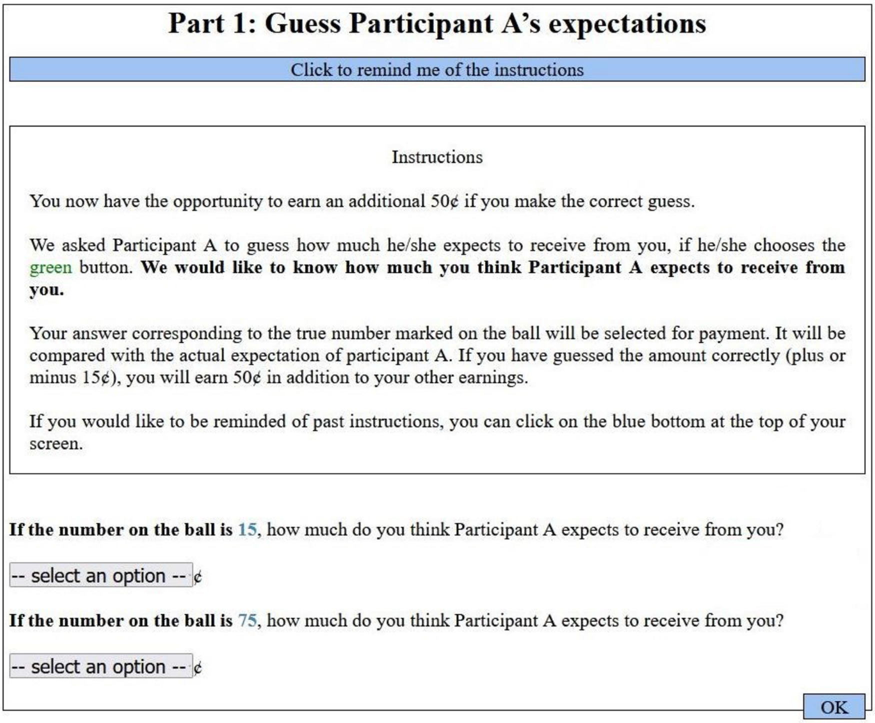
If you have any questions concerning this study, you can contact us at rimbaud@gate.cnrs.fr

Your confirmation code to be entered on Mturk webpage is your Mturk worker ID. Please submit the HIT on Mturk with this ID.

You can close this window now.

D.3 Trustees' screens in "Low Demand" Instructions

In the screen "Part 1: Guess Participant A's expectations", the two returns decisions are now made sequentially and in a random order: some participants first made their "Decision 15", then another screen their "Decision 75"; others faced the reverse order. We also removed the wording the reminder of their own guess "you reported that Participant A expects to receive x cents".



Part 1: Guess Participant A's expectations

Instructions

You now have the opportunity to earn an additional 50¢ if you make the correct guess.

We asked Participant A to guess how much he/she expects to receive from you, if he/she chooses the green button. **We would like to know how much you think Participant A expects to receive from you.**

Your answer corresponding to the true number marked on the ball will be selected for payment. It will be compared with the actual expectation of participant A. If you have guessed the amount correctly (plus or minus 15¢), you will earn 50¢ in addition to your other earnings.

If you would like to be reminded of past instructions, you can click on the blue bottom at the top of your screen.

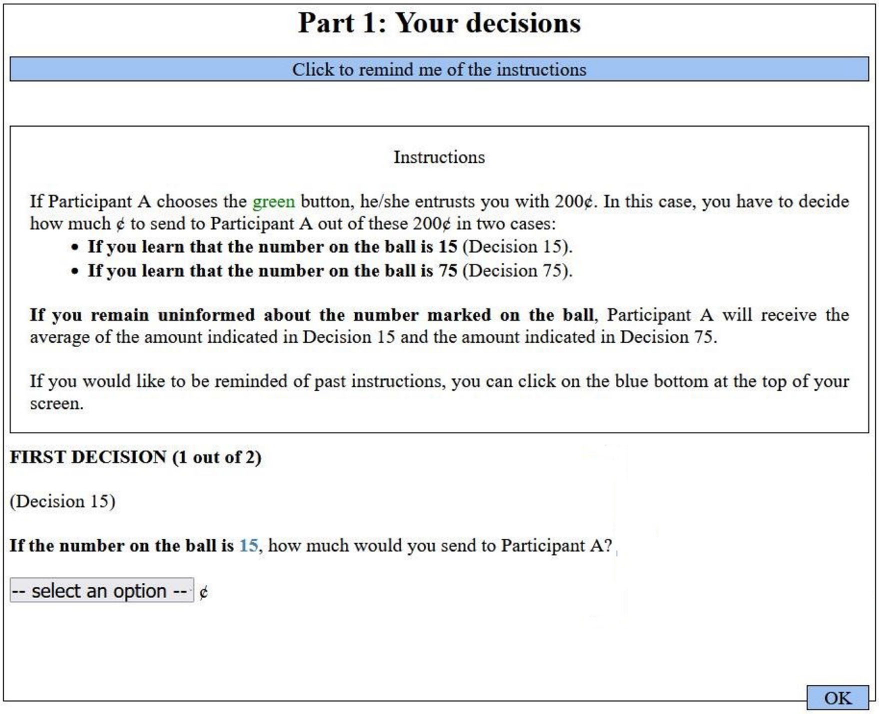
If the number on the ball is 15, how much do you think Participant A expects to receive from you?

-- select an option -- ¢

If the number on the ball is 75, how much do you think Participant A expects to receive from you?

-- select an option -- ¢

In the screen "Guess Participant's A expectations", the two elicited beliefs are also presented in random order, the wording "participant A has to give up x cents" and the sentence "Please pay attention to the different scenarios to answer the following question" were also removed.



Part 1: Your decisions

Instructions

If Participant A chooses the green button, he/she entrusts you with 200¢. In this case, you have to decide how much ¢ to send to Participant A out of these 200¢ in two cases:

- **If you learn that the number on the ball is 15** (Decision 15).
- **If you learn that the number on the ball is 75** (Decision 75).

If you remain uninformed about the number marked on the ball, Participant A will receive the average of the amount indicated in Decision 15 and the amount indicated in Decision 75.

If you would like to be reminded of past instructions, you can click on the blue bottom at the top of your screen.

FIRST DECISION (1 out of 2)

(Decision 15)

If the number on the ball is 15, how much would you send to Participant A?

-- select an option -- ¢

Acknowledgements We are grateful to Giuseppe Attanasi, Laurent Denant-Boëmont, Si Chen, Christoph Vanberg, and Marie Claire Villeval for useful feedback. We thank Quentin Thevenet for his help in programming the experiment. Helpful comments were received from participants to the GATE-Lab and AWI internal seminar series, the JMA Conference in Rennes, and the Lyon-Maastricht workshop. This study has been funded by 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 Agence Nationale de la Recherche (ANR) and by a the Young Scholar Research Funding of the FSP EPOS (Universty of Innsbruck). Claire Rimbaud was supported by the Austrian Science Fund (FWF, SFB F63).

References

Andreoni, J., & Sanchez, A. (2020). Fooling myself or fooling observers? Avoiding social pressures by manipulating perceptions of deservingness of others. *Economic Inquiry, 58*(1), 12–33.

Attanasi, G., Battigalli, P., Nagel, R., et al. (2022). Disclosure of belief-dependent preferences in the trust game. IGIER Working Paper Series, 506.

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

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

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

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

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

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

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

Charness, G., Oprea, R., & Yuksel, S. (2021). How do people choose between biased information sources? Evidence from a laboratory experiment. *Journal of the European Economic Association, 19*(3), 1656–1691.

Chen, S., Heese, C., et al. (2021). Fishing for good news: Motivated information acquisition. University of Bonn and University of Mannheim Discussion Paper Series, 223.

Chopra, F., Haaland, I., & Roth, C. (2023). The demand for news: Accuracy concerns versus belief confirmation motives. NHH Dept. of Economics Discussion Paper, 01.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. New York: Academic press.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory, 33*(1), 67–80.

Di Tella, R., Perez-Truglia, R., Babino, A., & Sigman, M. (2015). Conveniently upset: Avoiding altruism by distorting beliefs about others' altruism. *American Economic Review, 105*(11), 3416–42.

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology, 63*(4), 568.

Dufwenberg, M. (2002). Marital investments, time consistency and emotions. *Journal of Economic Behavior & Organization, 48*(1), 57–69.

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

Dufwenberg, M., Gächter, S., & Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior, 73*(2), 459–478.

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior, 47*(2), 268–298.

Exley, C. L. (2016). Excusing selfishness in charitable giving: The role of risk. *The Review of Economic Studies, 83*(2), 587–628.

Exley, C. L., & Kessler, J. B. (2021). Information avoidance and image concerns. Working Paper.

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

Feiler, L. (2014). Testing models of information avoidance with binary choice dictator games. *Journal of Economic Psychology, 45*, 253–267.

Festinger, L. (1962). Cognitive dissonance. *Scientific American, 207*(4), 93–106.

Fong, C. M., & Oberholzer-Gee, F. (2011). Truth in giving: Experimental evidence on the welfare effects of informed giving to the poor. *Journal of Public Economics, 95*(5–6), 436–444.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic behavior, 6*(3), 347–369.

Friedrichsen, J., Momsen, K., & Piasenti, S. (2022). Ignorance, intention and stochastic outcomes. *Journal of Behavioral and Experimental Economics, 100*, 101913.

Garcia, T., Massoni, S., & Villeval, M. C. (2020). Ambiguity and excuse-driven behavior in charitable giving. *European Economic Review, 124*, 103412.

Golman, R., Hagmann, D., & Loewenstein, G. (2017). Information avoidance. *Journal of Economic Literature, 55*(1), 96–135.

Grossman, Z., & Van Der Weele, J. J. (2017). Self-image and willful ignorance in social decisions. *Journal of the European Economic Association, 15*(1), 173–217.

Haisley, E. C., & Weber, R. A. (2010). Self-serving interpretations of ambiguity in other-regarding behavior. *Games and Economic Behavior, 68*(2), 614–625.

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

Hertwig, R., & Engel, C. (2016). Homo ignorans: Deliberately choosing not to know. *Perspectives on Psychological Science, 11*(3), 359–372.

Inderst, R., Khalmetski, K., & Ockenfels, A. (2019). Sharing guilt: How better access to information may backfire. *Management Science, 65*(7), 3322–3336.

Jia, T. (2021). Empathy, motivated reasoning, and redistribution. Working Paper.

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

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

Larson, T., & Capra, C. M. (2009). Exploiting moral wiggle room: Illusory preference for fairness? A comment. *Judgment and Decision Making, 4*(6), 467.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics, 1*(3), 593–622.

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal, 92*(368), 805–824.

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

Morell, A. (2019). The short arm of guilt: An experiment on group identity and guilt aversion. *Journal of Economic Behavior & Organization, 166*, 332–345.

Rabin, M. (1995). Moral preferences, moral constraints, and self-serving biases. *Berkeley Department of Economics Working Paper*, 95-241.

Saccardo, S., & Serra-Garcia, M. (2023). Enabling or limiting cognitive flexibility? Evidence of demand for moral commitment. *American Economic Review, 113*(2), 396–429.

Sengupta, A., & Vanberg, C. (2023). Promise keeping and reliance damage. *European Economic Review, 152*, 104344.

Serra-Garcia, M., & Szech, N. (2022). The (in)elasticity of moral ignorance. *Management Science, 68*(7), 4815–4834.

Smith, M. K., Trivers, R., & von Hippel, W. (2017). Self-deception facilitates interpersonal persuasion. *Journal of Economic Psychology, 63*, 93–101.

Soldà, A., Ke, C., Page, L., & Von Hippel, W. (2020). Strategically delusional. *Experimental Economics, 23*, 604–631.

Soraperra, I., van der Weele, J., Villeval, M. C., & Shalvi, S. (2023). The social construction of ignorance: Experimental evidence. *Games and Economic Behavior, 138*, 197–213.

Spiekermann, K., & Weiss, A. (2016). Objective and subjective compliance: A norm-based explanation of ‘moral wiggle room’. *Games and Economic Behavior, 96*, 170–183.

Woods, D., & Servátka, M. (2016). Testing psychological forward induction and the updating of beliefs in the lost wallet game. *Journal of Economic Psychology, 56*, 116–125.

Xiao, E., & Bicchieri, C. (2012). Words or deeds? Choosing what to know about others. *Synthese, 187*(1), 49–63.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics, 13*, 75–98.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.