No. 61

U N I V E R S I T Y O F C O L O G N E

W O R K I N G P A P E R S E R I E S I N E C O N O M I C S

SURPRISING GIFTS - THEORY AND LABORATORY EVIDENCE

KIRYL KHALMETSKI
AXEL OCKENFELS
PETER WERNER

Department of Economics
University of Cologne
Albertus-Magnus-Platz
D-50923 Köln
Germany

http://www.wiso.uni-koeln.de

Surprising Gifts

Theory and Laboratory Evidence

Kiryl Khalmetski,[1] Axel Ockenfels,[2,*] Peter Werner[3]

9 May 2013

Abstract

People do not only feel guilt from not living up to others' expectations (Battigalli and Dufwenberg (2007)), but may also like to *exceed* them. We propose a model that generalizes the guilt aversion model to capture the possibility of positive surprises when making gifts. A model extension allows decision makers to care about others' attribution of intentions behind surprises. We test the model in two dictator game experiments. Experiment 1 shows a strong causal effect of recipients' expectations on dictators' transfers. Moreover, in line with our model, the correlation between transfers and expectations can be both, positive and negative, obscuring the effect in the aggregate. Experiment 2 shows that dictators care about what recipients know about the intentions behind surprises.

Keywords: guilt aversion, surprise seeking, dictator game, consensus effect

JEL-Codes: C91, D64

Financial support of the German Science Foundation (DFG) through the Leibniz program and through the Research Unit "Design & Behavior" (FOR 1371) is gratefully acknowledged. We thank Martin Dufwenberg, Roman Inderst, Jean-Robert Tyran and participants at the ESA European Conference 2012 in Cologne for helpful comments and suggestions.

[*] Corresponding author.

[1] University of Frankfurt, Senckenberganlage 31, D-60325 Frankfurt am Main, Germany (e-mail: kiryl.khalmetski at hof.uni-frankfurt.de).

[2] University of Cologne, Albertus-Magnus-Platz, D-50923 Köln, Germany (e-mail: ockenfels at uni-koeln.de).

[3] University of Cologne, Albertus-Magnus-Platz, D-50923 Köln, Germany (e-mail: peter.werner at uni-koeln.de).

1. INTRODUCTION

Models of guilt aversion assume that people feel guilt from not living up to others' expectations (Battigalli and Dufwenberg (2007), BD in the following). Yet, it appears plausible that some people do not only suffer from negative surprises, but may also get pleasure from positive surprises (Mellers et al. (1997)). We thus generalize the model of guilt aversion by incorporating the notion that people may care for both, positive and negative surprises when making gifts.[1] We test the model's predictions in two dictator game experiments and find strong support. Moreover, we show that our data reconcile seemingly conflicting evidence from previous studies on guilt aversion.

Our Experiment 1 is designed to investigate the prediction that dictator transfers can both, decrease and increase with the recipient's expectation, depending on the weight put on positive and negative surprises, respectively. We find a strong causal effect of recipients' expectations on individual dictator transfers. The effect is obscured on the aggregate level because, as suggested by our model, dictators differ in how they react to the recipients' expectations.

Other studies, too, found that others' expectations may directly affect social behavior. Eliciting subjects' beliefs about the expectations of interaction partners (second-order beliefs, SOBs), several studies detected a positive relation between beliefs and observed behavior. In one study by Charness and Dufwenberg (2006), subjects who held significantly higher beliefs about their transaction partner's expectation were also more trustworthy. Several other experiments have found positive correlations between subjects' self-reported beliefs and observed decisions.[2]

[1] Our research is part of the literature that is devoted to people's concern about beliefs *per se*, independently of the material outcome (Geanakoplos et al. (1989), Bénabou and Tirole (2006), Andreoni and Bernheim (2009)). The framework of dynamic psychological games (Battigalli and Dufwenberg (2009)) incorporates many of these earlier approaches, including the notion that people suffer from guilt when they disappoint what they think are other players' expectations.

[2] For experimental evidence on the impact of belief-dependent preferences in trust, dilemma and principal-agent games see also Guerra and Zizzo (2004), Falk and Kosfeld (2006), Dufwenberg et al. (2011) and Charness and Dufwenberg (2011). Vanberg (2008) investigated potential reasons behind the positive effect of promises on trustworthy behavior found in Charness and Dufwenberg (2006) and concluded that preferences for promise-keeping rather than preferences for meeting expectations might be the predominant driver of the results. Also, with respect to dictator games, the willingness of some dictators to exploit information asymmetries between themselves and

However, more recently, some authors have argued that correlations between self-reported SOBs and choices may be confounded by the false consensus effect (Ross et al. (1977)): the SOB might be biased towards one's own choice. If this is the case, observed correlations between actions and beliefs are no conclusive evidence for beliefs causally affecting behavior.[3]

To address concerns about consensus effects, Ellingsen et al. (2010, EJTT in the following) *induced* SOBs in experimental dictator games by disclosing the first-order beliefs (FOBs) elicited from recipients to dictators. With this design, the authors made it possible to establish a direct *causal* influence of SOBs on dictator transfers. Yet, no correlation was found between induced SOBs and actual transfers, leading the authors to the conclusion that the empirical relevance of guilt aversion might be limited and partly confounded by the false consensus effect. Our experiments closely follow the design by EJTT and also induce SOBs. This way, we replicated all main results of EJTT's dictator treatment, including the lack of correlation between transfers and induced SOBs *on the between-subject level*. Moreover, we provide further evidence for the confounding role of the false consensus effect. However, in addition to EJTT, we utilize the strategy method (Selten (1967)), eliciting transfers for all possible expectation levels of the recipient. This aspect of our design allows us to investigate the different individual patterns of behavior that we expected to see based on our model. The *within-subject* data show that most subjects do systematically condition transfers on the recipients' expectations. Yet, because we observe both, positive and negative within-subject correlations of transfers with expectations, no such correlation can be identified for the aggregated data.

In Experiment 2, we take a complementary approach to study the performance of our model in the laboratory. Here, we are not interested in identifying individual differences, but in designing a situation, where the comparative static prediction of our model is the same for *all* preferences allowed by the model. At the same time, Experiment 2 investigates in more detail the nature of the dictator's motivation for surprising. More specifically, BD introduced two models of guilt aversion. One is

recipients suggests that altruistic behavior may depend on beliefs (see, for example, Dana et al. (2007), Andreoni and Bernheim (2009), Grossman (2010), Ockenfels and Werner (2012)).

[3] Selten and Ockenfels (1998), for instance, found evidence for strong consensus effects in the solidarity game, a variant of the dictator game.

simple guilt and refers to a player who cares about the extent to which he lets another player down. The second assumes that a player cares about others' inferences regarding the extent to which he is willing to let them down (i.e. inferences about his intentions). We formulate our model to capture the potential role of 'intentional surprise'. The model predicts that if the recipient's inference about the dictator's intention is ambiguous, the latter has weaker incentives both to avoid guilt and to positively surprise the recipient, and should in turn transfer less. The effect is predicted for both relatively guilt-averse and surprise-seeking dictators. To test this prediction we introduce an experimental design, which manipulates the recipient's inference about the dictator's intentions through making the recipient either aware or unaware about the fact that the dictator's SOB was induced. The results of the experiment support our model's prediction, and show that dictators care about the recipients' attribution of their intentions behind surprises.

Section 2 presents our generalized model of surprising, describes the experimental design to investigate both, guilt aversion and surprise seeking, analyzes the data and compares them to related results in the literature. Section 3 extends the model of surprising to capture the effect of the recipient's inference about the dictator's intention, and presents the design and the results of the second experiment. Section 4 concludes.

2. A MODEL OF SURPRISING OTHERS AND EXPERIMENT 1

2.1. Model

Assume that dictator *i* divides an amount normalized to 1 between himself and recipient *j*, who holds an ex ante expectation about her dictator's transfer t_i. Applying the benchmark model of guilt aversion of BD, the expected utility of the dictator is given by

$$U_i(t_i, E_{ij}) = m_i(1-t_i) - \lambda_i G_i(t_i, E_{ij}), \quad (1)$$

where $m_i(\cdot)$ is the standard utility of money, further assumed to have conventional properties $m_i'(\cdot) > 0,\ m_i''(\cdot) \leq 0$, $E_{ij} = E_i E_j[m_j(t_i)]$ is the recipient's expectation about her monetary utility as expected by the dictator (i.e. the SOB of the dictator), and $G_i(t_i, E_{ij}) = \max[0, E_{ij} - m_j(t_i)]$ is the level of guilt from falling below the recipient's expectation.

In the formalization by BD, guilt is strictly positive only for transfers strictly below expectations. That is, only negative surprises matter. However, based on the idea that people like pleasant surprises, it appears reasonable that both, negative and positive deviations from the recipient's expectation directly enter the dictator's utility function. More specifically, we assume that dictators do not only suffer from negatively surprising the recipient, but also derive utility from positively surprising her.

Moreover, while models of guilt aversion take as the recipient's reference point her point expectation, some other models of reference-dependent preferences (like Köszegi and Rabin (2006)) exploit the whole distribution of beliefs (a reference lottery). That is, the ex post outcome is compared with all outcomes in the support of the reference lottery weighted by the corresponding ex ante probabilities. Following this approach, we further deviate from guilt aversion models by assuming that the reference point of the recipient, against which the surprise is evaluated, is given by a probability distribution of possible outcomes (i.e. the reference point is stochastic). We further denote the cumulative distribution function (cdf) of the FOB of the recipient as H_j, with the corresponding probability density function (pdf) h_j. Correspondingly, the SOB of the dictator is given by cdf $H_{ij}(x) = E_i[H_j(x)]$, with pdf denoted by $h_{ij}(x)$.[4] These assumptions lead to the following extension of the dictator's expected utility function (1):

[4] While BD's original approach based on the point-wise representation of beliefs has many benefits due to its simplicity, it implies that the marginal effect of the SOB on the sender's utility (given linearity of the guilt function) is constant. This limits the analysis of the comparative statics of the sender's optimal transfer with respect to the SOB, which is the focus of the subsequent study in this section. As we show later, the distribution-wise representation of beliefs provides a natural way to make the optimal transfer monotonically changing with the SOB, which enhances the predictive power of the model. An alternative approach could be to introduce some nonlinearity directly into the guilt

$$U_i(t_i \mid h_{ij}) = m_i(1-t_i) + S_i(t_i \mid h_{ij}), \tag{2}$$

with

$$S_i(t_i \mid h_{ij}) = \alpha_i \int_0^{t_i} (t_i - x) h_{ij}(x) dx - \beta_i \int_{t_i}^1 (x - t_i) h_{ij}(x) dx \,. \tag{3}$$

Hereafter, $S_i(t_i \mid h_{ij})$ is referred to as the surprise function. The first term in the surprise function represents the dictator's expected utility from positive surprises (when $x < t_i$), while the second one the expected disutility from negative surprises (when $x > t_i$). The (stochastic) reference point is the distribution of SOB, given by the pdf $h_{ij}(x)$. Correspondingly, the scalar $\alpha_i \geq 0$ denotes the propensity to make positive surprises (surprise seeking), and the scalar $\beta_i \geq 0$ corresponds to the propensity to avoid negative surprises (guilt aversion). These propensities are not necessarily equal. Note that, in order to simplify the exposition, we assume that the value of surprise for a particular belief x and transfer t_i (the term weighted by $h_{ij}(x)$) is linear in $x - t_i$.[5]

In what follows, we make the following assumptions on the utility function:

A1. $\alpha_i + \beta_i > 0$.

A2. $m_i'(1-t_i) \geq \dfrac{\alpha_i + \beta_i}{2}$ for any $t_i \in [0,1]$.

Assumption A1 simply implies that the dictator has some preferences regarding (either positively or negatively) surprising the recipient.[6]

---

function, which we find less plausible due to (at least to some degree) arbitrary nature of such functional restrictions.

[5] The original negative surprise term in the utility function (1) is not necessarily linear. Our simplification is not critical for our Proposition 2 in this section, but we need it for our Propositions 1 and 3.

[6] When both coefficients are 0, the dictator chooses zero transfers for any beliefs. In principle, one could also add other motives, not related to beliefs, such as inequality aversion to the utility function. Then, there can be positive equilibrium transfers even if $\alpha_i = \beta_i = 0$. Yet, the transfers will still be independent of beliefs.

Assumption A2 states that the marginal monetary cost of giving is larger than the average sensitivity to positive and negative surprises. For example, if $\alpha_i = 0$, $\beta_i > 0$ and the utility of money is linear, then Assumption A2 requires that for an expected decrease in the negative surprise term by 1 Euro a dictator is willing to pay at most 2 Euro.[7] The assumption is well in line with existing estimations of the quantitative effect of guilt aversion (Bellemare et al. (2011)).

Regarding the information structure of the game, we assume that the FOB of the recipient is unknown to the dictator. Yet, he observes an informative signal θ_j about the FOB, which is equal to the median of the FOB distribution. Then, his SOB is characterized by the conditional cdf $H_{ij}(x \mid \theta_j) = E_i[H_j(x) \mid \theta_j]$ with the corresponding conditional pdf $h_{ij}(x \mid \theta_j)$. It holds that θ_j is the median of the dictator's conditional SOB distribution as well:

$$H_{ij}(\theta_j \mid \theta_j) = E_i[H_j(\theta_j) \mid \theta_j] = E_i[\frac{1}{2}] = \frac{1}{2}. \tag{4}$$

We emphasize that we do not require that individual beliefs correspond to a rational expectations equilibrium. Rather, we treat θ_j as exogenous to the model. Mutual consistency of beliefs and behavior is rejected by numerous dictator game experiments, including EJTT (see their Figure 1, which reveals significant heterogeneity in beliefs about average dictator transfers). Of course, our modeling does neither exclude the possibility that beliefs are consistent with behavior, nor that average beliefs are consistent with average behavior (as roughly observed by Selten and Ockenfels (1998), among others).

Further, we do not explicitly model how the dictator forms his SOB as the expectation of the recipient's FOB conditional on the obtained signal. Instead, we implement a reduced-form model, assuming only that a higher signal leads to a higher SOB in the sense of first-order stochastic dominance (FOSD).

[7] In this case the decrease in the negative surprise term by 1 Euro yields an increase in the total utility equal to $\beta_i \Delta$, and the payment of additional 2 Euro results in a loss of $m_i'(1-t_i)2\Delta$, where Δ corresponds to 1 Euro in terms of normalized amounts. Assumption A2 implies that the loss is weakly higher than the gain.

A3. The SOB conditional on a higher signal (strictly) first-order stochastically dominates the SOB conditional on a lower signal:

$H_{ij}(x \mid \theta_j'') < H_{ij}(x \mid \theta_j')$ if and only if $\theta_j'' > \theta_j'$ for any $x \in (0,1)$ and $\theta_j', \theta_j'' \in [0,1]$.[8]

Finally, we impose some smoothness on the cdf function $H_{ij}(x \mid \theta_j)$:

A4. $H_{ij}(x \mid \theta_j)$ is continuously differentiable on $[0,1]\times[0,1]$.

Assumption A4 implies that Assumption A3 can be reformulated as

$$\frac{\partial H_{ij}(x \mid \theta_j)}{\partial \theta_j} < 0 \text{ for any } x \in (0,1) \text{ and } \theta_j \in [0,1]. \tag{5}$$

For ease of notation let us further denote the surprise and the utility functions given signal θ_j as $S_i(t_i, \theta_j)$ and $U_i(t_i, \theta_j)$, respectively.

Let us now consider the optimal strategy of the dictator. For simplicity, we assume that the dictator plays a pure strategy conditional on the signal θ_j, i.e. the dictator's chosen transfer can be represented as a function $t_i^*(\theta_j)$ of the signal, so that[9]

$$t_i^*(\theta_j) \in \arg\max_{t_i} U(t_i, \theta_j). \tag{6}$$

First, we establish the result that if the dictator has relatively surprise seeking (guilt averse) preferences, then his optimal transfer exceeds (falls below) the signal about the recipient's expectation:

Proposition 1. *The optimal transfer $t_i^*(\theta_j)$ is weakly larger than θ_j if $\alpha_i > \beta_i$, and weakly smaller than θ_j if $\beta_i > \alpha_i$, provided that $t_i^*(\theta_j)$ is non-zero.*

[8] $H_{ij}(x \mid \theta_j)$ does not depend on θ_j at $x = 0$ and $x = 1$ (being always equal to 0 and 1, respectively).

[9] We do not exclude that the dictator can be indifferent between multiple transfers for some θ_j.

Proof. See Appendix A. ■

Intuitively, for a surprise-seeking dictator the mass of beliefs affected by positive surprising (i.e. below the transfer) should be sufficiently large to compensate the monetary cost of giving. Hence, the transfer should itself be relatively large (that is, above the median θ_j). The opposite logic applies for relatively guilt-averse dictators.[10]

Proposition 2 shows how the optimal transfer $t_i^*(\theta_j)$ depends on the dictator's signal θ_j about the recipient's FOB.[11]

Proposition 2. *$t_i^*(\theta_j)$ is increasing (decreasing) in θ_j if $\alpha_i < \beta_i$ ($\alpha_i > \beta_i$), being strictly increasing (decreasing) if $0 < t_i^*(\theta_j) < 1$.*

Proof. See Appendix A. ■

The result is driven by the fact that a higher signal θ_j implies by Assumption A3 that for each transfer a larger probability mass is placed on the recipient's belief above the transfer, which is affected by negative surprising. In contrast, a lower probability mass is placed on the belief below the transfer, which is subject to positive surprising. This results in a higher marginal cost of negative surprising and a lower marginal gain from positive surprising. Consequently, an optimal transfer is pushed upward from the perspective of negative surprises, but downward from the perspective of positive surprises. Which force prevails depends on the relationship between the dictator's relative preference for positive and negative surprises as denoted by α_i and β_i. For example, if a dictator cares more about avoiding negative surprises $(\alpha_i < \beta_i)$, the optimal transfer will increase with the received signal about the recipient's FOB.

[10] Any relative preferences can support an optimal transfer of 0, when the belief-dependent preferences are negligible relative to the monetary utility. If $\alpha_i = \beta_i$, then our assumptions do not preclude the optimal transfer to be at any value (see Appendix A).

[11] The proof of Proposition 2 does not require Assumption A2 and the fact that θ_j is the median of the recipient's FOB distribution, but is consistent with them. For the case $\alpha_i = \beta_i$ one can show that the set $\arg\max_{t_i} U(t_i, \theta_j)$ does not depend on θ_j.

2.2. Experiment design and hypotheses

Our Experiment 1 is designed to test whether surprise seeking contributes to explaining dictator game behavior. We decided to choose an experimental setting that closely follows the design of the dictator game experiment by EJTT, who informed dictators about recipients' expectations before the transfer had to be chosen. The only exception is that we used the strategy method in our experiment in order to elicit dictator transfers *conditional* on the recipient's potential expectations, whereas EJTT employed the play-method. One reason for using EJTT's paradigm is that EJTT could not find evidence for guilt aversion in their data. Since our model is built on guilt aversion as a major driver of behavior, the EJTT design thus constitutes a particularly challenging test for our model. We choose the strategy method because, as we show below, this allows us not only to replicate the EJTT findings (as a robustness check), but also to detect a potential heterogeneity in dictators' transfers as a function of expectations, and so to provide a new interpretation of EJTT's dictator game data.

In our experiment, each dictator had to divide 14 Euro between himself and a randomly matched anonymous recipient. Before observing the actual amount sent to her, the recipient was asked to provide a guess for the average transfer in the population.[12] Before the guess of the recipient was revealed to the corresponding dictator, he was asked to indicate his transfer conditional on all possible guesses rounded to 50 cents. Guesses higher than 9 Euro were grouped into a single category. After that, the conditional transfer, which corresponded to the actual expectation level of the recipient, was implemented and paid out.[13]

The experiment was conducted as a classroom experiment among students of economics and business at the University of Frankfurt with a total of 386 students.

[12] In order to stay close to EJTT's design, the guess closest to the average transfer was rewarded with an additional bonus of 8 Euro.

[13] In line with the design of EJTT, at the time of guessing, recipients did not know that dictators can make their transfer conditional on the guess. EJTT discuss in their paper why not telling subjects about the guesses being communicated to dictators does not generally violate the no-deception norm in experimental economics. One might argue, however, that dictator behavior might get distorted, because they might get suspicious when learning, before making their choices, that recipients are not fully informed about the rules of the game being played. However, we have no indication for such distortions. Our data is fully consistent with related dictator game data. Moreover, there is no reason to suppose that such a suspicion would translate into behavior that responds to the comparative statics as observed in our Experiments 1 and 2 and as predicted by our model.

The classroom was divided in two separate halves by a central aisle. Students sitting in the first half received the dictator's instructions, and students sitting in the second one received the recipient's instructions. Instructions can be found in Appendix D.

The recipient's guess can be plausibly assumed to be close to the median of her FOB distribution, and hence we assume that it corresponds to signal θ_j in terms of our model. Then, the strategy method allows us to infer the mapping of signals θ_j to transfers $t_i^*(\theta_j)$ for each dictator. In other words, the design allows us to investigate a dictator's willingness to give as a function of the recipient's expectation – in contrast to previous studies of guilt aversion where the correlation between transfers and expectations could be observed only at the between-subject level. Our null hypothesis is that the dictators do not care about either positively or negatively surprising the recipients. This means that strictly positive transfers are driven only by outcome-based social preferences, such as narrowly selfish preferences or inequality aversion (Fehr and Schmidt (1999) and Bolton and Ockenfels (2000)). If this is the case, transfers are fully independent of the recipients' guesses. However, if the dictator (additionally) cares about the recipient's expectation, Propositions 1 and 2 suggest that we should observe transfers both below and above the expectation *and* that those transfers are systematically (positively or negatively) correlated with expectations.

2.3. Experimental results

In total we obtained 3,629 observations for conditional transfers from 191 dictators (19 for each dictator), and 195 observations for recipients' guesses.[14] Our results are comparable to the results in the dictator game treatment of EJTT. The average actually realized transfer is 3.25 Euro. This is 23% of the endowment, which is approximately the same value as in EJTT, where it was 24% of the endowment.

[14] Because of a matching error, we had four recipients more than needed; these recipients were paid according to the decisions made by a randomly chosen dictator of another pair. A single dictator provided only one conditional transfer, while leaving the fields for other conditional transfers blank. These blank fields were interpreted as zeros, though, none of our results are affected if we drop this dictator. Also, we always include transfers conditional on the last top-coded guess in the data analyses, though, our conclusions do not change if we exclude these values.

28% of the dictators are not willing to transfer a strictly positive amount to their matched recipients. The corresponding value is 35% in EJTT. The average guess of recipients is 4.70 Euro. This is 34% of the endowment, compared to 32% in EJTT. Finally, EJTT emphasized that they did not find a correlation between guesses and transfers (Pearson correlation coefficient of −0.075, $p = 0.497$). In our experiment, the correlation between actually realized transfers and guesses, too, is not significantly different from zero (Pearson correlation coefficient of −0.017, $p = 0.821$).[15]

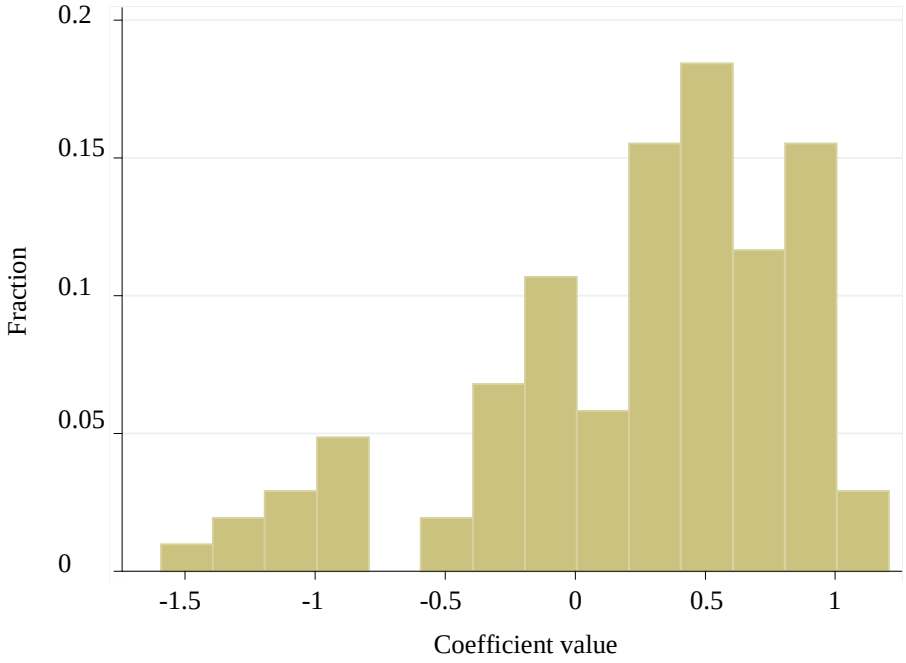
All these observations are in line with our null hypothesis. However, the within-subject data tell a more subtle story: 77% of the dictators change their transfers in response to guesses at least once, and 54% of the dictators exhibit a within-subject correlation of transfers with guesses which is significant at the 5% level. To check whether the observed pattern can be organized by a random process, we ran a Monte-Carlo simulation with 10,000 replications of random samples of transfers obtained by bootstrapping the original sample. On average, the share of significant within-subject correlations between transfers and guesses in random samples is just 3.7% with a standard deviation of 1.3%. None of the replications produced a sample with a share of significant correlations of more than 9.4%. We conclude that our observed share of 54% is the result of a systematic choice. This rejects all purely outcome-based models as an explanation of positive transfers, and demonstrates that many dictators care about recipients' beliefs.

Consistent with Proposition 2, we also find that dictators differ qualitatively in how they respond to changes in recipients' expectations. Figure 1 shows the distribution of the statistically significant coefficients from regressing transfers on guesses for each dictator. According to Proposition 2 positive regression coefficients correspond to relatively guilt-averse dictators, while negative coefficients to relatively surprise-seeking dictators. Figure 1 shows that 70% of the coefficients are distributed to the right of zero. The asymmetry is statistically significant: a two-sided sign test

[15] Random matching is itself a stochastic parameter, and hence a single random matching may be not representative. Our within-subject design allows a more robust measure of the average correlation at the between-subject level by estimating correlation coefficients under different possible matchings between dictators and recipients. We performed a Monte-Carlo simulation of 10,000 random matching combinations between subjects, estimating correlation between transfers and guesses in each replication. The average Pearson correlation coefficient is with 0.102 ($p = 0.162$) a bit higher than the low coefficient that corresponds to the random matching used to pay out our subjects, but it is still not significantly different from zero.

strongly rejects the hypothesis that the median is equal to 0 ($p < 0.001$). Also, the average size of positive coefficients (0.58) is somewhat larger than the size of negative coefficients (0.53), although the difference in distributions is not statistically significant ($p = 0.123$, two-sided MWU-test). Hence, guilt aversion appears to be more prevalent than surprise seeking in our dictator game context. This seems consistent with reference-dependent preferences models (like Kahneman and Tversky (1979), Fehr and Schmidt (1999), Köszegi and Rabin (2006)) along with empirical evidence (like Tversky and Kahneman (1992), Fehr and Schmidt (1999)), which suggests that falling below the respective reference standard generally has a larger effect on utility than a same-sized gain above the reference point.

Fig. 1. The distribution of coefficients, significant at 5% level, estimated in within-subject regressions of transfers on guesses.



At the same time, we find that the positive surprise side cannot be neglected either. For one, 27% of all transfers submitted by dictators are strictly above guesses.[16] Second, and more importantly, 30% of the dictators for whom we find a significant within-subject correlation between transfers and guesses exhibit a *negative*

[16] EJTT also found a significant share of such transfers (see their Figure 1).

correlation. This corresponds to 16% of the total population.[17] Both observations are inconsistent with *pure* guilt aversion, yet consistent with our generalized model (Proposition 2).[18]

To further investigate the nature of the negative correlation in our data we analyze whether this correlation is more typical for relatively surprise-seeking subjects, as our Proposition 2 suggests. Proposition 1 establishes that surprise-seeking subjects should exceed their recipients' guesses (in contrast to guilt-averse dictators). Hence, to test our theory we check whether transfers above guesses are negatively correlated with guesses, and transfers below guesses exhibit a positive correlation. A test that simply looks at transfer observations above or below corresponding guesses, respectively, would be distorted, though. The reason is that such subsamples of transfers are positively correlated with guesses by construction (i.e., even in random samples), since the midpoint of the range of selected transfers is then increasing with the guess. This is why we estimate truncated regression coefficients (controlling for fixed effects at the individual level), truncating the total sample of conditional transfers from below at different thresholds and starting at the threshold of 1 Euro. These coefficients reflect the dictators' sensitivity towards recipients' expectations in subsamples of transfers above corresponding thresholds.[19] Under the null hypothesis that the correlation between transfers and guesses is independent of whether the transfer is above or below the guess, the coefficients should not differ by thresholds. The alternative hypothesis is that the coefficients decline with the threshold as the relative share of the positive surprise segment (where $t_i^*(\theta_j) > \theta_j$) is increasing.

[17] The highest share of dictators with a significantly negative within-subject correlation observed in 10,000 bootstrapped replications of random transfers was 6.3% (with 95% of replications yielding a share below 3.7%). Hence, the share of negative correlations of 16% is outside any bootstrap confidence interval.

[18] One might argue that the strategy method induces a demand effect: subjects may consider it 'appropriate' to take the others' expectations into account, because they are given an explicit choice to condition their decisions on expectations. However, observe that our results, to the extent comparable, are fully consistent with EJTT's results, who do not employ the strategy method. Also, as a further robustness check, observe that our Experiment 2 provides clean and strong evidence for the role of the recipients' expectations in a play-method experiment. Finally, several studies report no evidence for the strategy method to yield qualitatively different results, like Brandts and Charness (2011) and Fischbacher et al. (2012).

[19] Regarding potential biases in the estimation of truncated regression models with fixed effects, Greene (2004) conducted a simulation study of several maximum likelihood estimation techniques and found no substantial biases in the truncated regression model.

Table 1 shows that the coefficient on guesses is gradually turning from significantly positive to significantly negative as the selected range of transfers shrinks towards the top.[20] This supports our modeling approach that different dictators put different weight on their willingness to respectively positively and negatively surprise others.

Table 1.
Fixed-effects truncated regression coefficients for subsamples of conditional transfers.

Range of transfers	Regression coefficient	p-value	Number of observations	Number of subjects
$t > 1$	0.181	0.000	2045	148
$t > 2$	0.123	0.000	1772	140
$t > 3$	0.045	0.053	1483	132
$t > 4$	-0.042	0.086	1136	117
$t > 5$	-0.104	0.000	889	101
$t > 6$	-0.151	0.000	692	93
$t > 7$	-0.256	0.001	200	33
$t > 8$	-0.197	0.017	139	28
$t > 9$	-0.141	0.041	108	21
$t > 10$	-0.204	0.001	63	12
$t > 11$	-0.096	0.008	53	12

The coefficients are obtained from regressing the truncated sample of all conditional transfers revealed by strategy method on guesses, under different truncation thresholds, no control variables included. P-values are obtained from robust standard errors.

2.4. False consensus

EJTT interpret the fact that there is no correlation between transfers and *induced* SOBs suggesting "*that consensus effects are responsible for a substantial fraction of the correlation between second-order beliefs and behavior in other studies*" (p. 101). In this view, the correlation is not due to SOBs causally affecting behavior, but rather due to a tendency of subjects to believe that others' behavior is similar to one's own behavior. However, our model suggests and our data show that EJTT's non-

[20] The results of Table 1 take into account within-subject variability of transfers, as they are based upon individual strategy vectors of dictators. Therefore, the results are not directly comparable to the between-subject analysis provided before based on the realized matchings of dictators and recipients.

correlation result is caused by opposing causal effects of SOBs across individual dictators – guilt aversion and positive surprise seeking.[21]

That said, our data confirm EJTT's conjecture and others' findings that false consensus is present in social contexts. We measured the ex ante SOBs of dictators in a survey, which the dictators had to complete *after* the transfer decisions have been made but *before* knowing the true guess of their recipient. The survey asked to provide an estimate of the average recipient's guess. The correlation between these self-reported SOBs and the corresponding conditional transfers is highly significant with a coefficient of 0.438 ($p < 0.001$). That is, if transfers were chosen according to the self-reported SOB, expectations and transfers would have been strongly correlated. We also observe that the absolute difference between transfers and SOBs is significantly smaller for the self-reported SOB (2.06 on average) compared to the induced SOB (3.20 on average; $p < 0.001$, two-sided sign test).[22] Overall, we conclude that the false consensus effect is strong and likely contributes to the observed significant effect of SOBs in studies based on self-reported SOB. The fact that transfers are relatively close to self-reported SOB is neither predicted by, nor inconsistent with our model. However, the false consensus effect does not organize how dictators respond to induced SOBs, which is the focus of our model and experiments.

3. A MODEL OF INTENTIONAL SURPRISE AND EXPERIMENT 2

The goal of Experiment 1 was to establish that there is a lot of (systematic) heterogeneity regarding how dictators respond to recipients' expectations. The goal of Experiment 2 is to show that, although there is much heterogeneity, there are settings

[21] One way to contrast the insignificant overall correlation with the significant within-subject correlation is to correlate guesses larger than zero with the *absolute* value of the difference between the transfer at those guesses and the transfer chosen for a guess equal to zero (i.e. with the absolute change in transfers). If transfers were just chosen randomly, we would expect a zero correlation between guesses and absolute changes in transfers. Yet, this correlation is highly significant with a coefficient of 0.217 ($p = 0.004$). Performing a Monte-Carlo simulation with different dictator-recipient matching combinations as a robustness check leads to an average correlation coefficient of 0.223 ($p = 0.003$).Thus, the absolute effect of guesses on transfers, which neutralizes the different sign of the effect across subjects, is highly statistically significant.

[22] The result is robust to rematching of subjects, as confirmed by Monte-Carlo simulations.

in which the comparative statics of incentives is perfectly aligned for all dictators, regardless of whether they care more about negative or positive surprises. At the same time, this section demonstrates that our model can easily be extended to also capture that dictators may care of the attribution of intentionality, much like the "guilt from blame" model introduced by BD. Specifically, our extended model predicts that if the recipient's inference about the dictator's intention is ambiguous, transfers are smaller. This holds independently of how much weight is put on negative and positive surprises, respectively.

3.1. Model

We introduce a generalization to the expected utility function (2) of the dictator in order to capture the possibility that the dictator cares about the attribution of intentions:

$$U_i(t_i \mid h_{ij}, h_{ijij}) = m_i(1-t_i) + \lambda_1 S_i^S(t_i \mid h_{ij}) + \lambda_2 S_i^I(t_i \mid h_{ijij}), \tag{7}$$

where

$$S_i^S(t_i \mid h_{ij}) = \alpha_i \int_0^{t_i} (t_i - x) h_{ij}(x)dx - \beta_i \int_{t_i}^1 (x - t_i) h_{ij}(x)dx, \tag{8}$$

$$S_i^I(t_i \mid h_{ijij}) = \alpha_i \int_0^{t_i} (t_i - x) h_{ijij}(x \mid t_i)dx - \beta_i \int_{t_i}^1 (x - t_i) h_{ijij}(x \mid t_i)dx. \tag{9}$$

Here S_i^S coincides with surprise function S_i considered in the previous section: it denotes utility derived from a *simple* surprise that the dictator experiences directly when deviating from the recipient's expectation. In contrast, S_i^I denotes utility derived from the recipient's attribution of the *intentions* behind the surprise (which we refer to below as 'intentional surprise'). The only difference between S_i^I and S_i^S is that the SOB density $h_{ij}(\cdot)$ is replaced by the (posterior) fourth-order belief density $h_{ijij}(\cdot \mid t_i)$ conditional on transfer t_i, with corresponding cdf $H_{ijij}(\cdot \mid t_i)$. The latter is

constructed as follows: the recipient's third-order belief $H_{jij}(x \mid t_i)$ represents her inference about the actual SOB of the dictator conditional on observing transfer t_i, that is $H_{jij}(x \mid t_i) = E_j[H_{ij}(x) \mid t_i]$. In turn, the dictator's fourth order belief is the inference of the dictator about the recipient's conditional third-order belief: $H_{ijij}(x \mid t_i) = E_i[H_{jij}(x \mid t_i)]$. Therefore, S_i^I corresponds to the expected recipient's inference about whether a deviation from expectation is by the dictator's intention to surprise the recipient, or due to his SOB (erroneously) deviating from the recipient's FOB. In the latter case, the surprise effect can be mitigated for the dictator, as we show below.

BD assume, in the analysis of their concept of guilt from blame, that the recipient blames the first player more if the negative surprise has been intentional (that is, expected by the first player). More blame, in turn, implies a larger utility loss for the first player. Analogously, we assume that an intentional positive surprise leads to more appreciation and gratitude from the recipient than a positive surprise that has occurred due to a dictator's confusion about the true expectations of the recipient. Such additional gratitude then leads to a larger utility gain for the dictator from positively surprising the recipient.

The coefficients $\lambda_1 \geq 0$ and $\lambda_2 > 0$ denote the relative weights of S_i^S and S_i^I, respectively, in the dictator's utility, such that their sum is normalized to 1:[23]

$$\lambda_1 + \lambda_2 = 1. \tag{10}$$

We keep Assumptions A1-A4 laid out in the last section.[24] At the same time, we introduce the following two information treatments, which will correspond to our laboratory treatments in Experiment 2: PUBLIC and PRIVATE. In the PRIVATE treatment the recipient remains unaware that the dictator observes θ_j before his

[23] We assume λ_2 to be strictly positive because we want to investigate the impact of intentional surprise S_i^I.

[24] That is, we assume that the dictator gets a signal θ_j about the median of the recipient's FOB, subject to Assumptions A3 and A4. For the subsequent results, it is sufficient that Assumption A2 holds only in the PUBLIC treatment (as explained below), which is strategically equivalent to the model setting in the previous section.

decision and, importantly, the dictator knows about this unawareness. In contrast, in the PUBLIC treatment the signaling of the recipient's ex ante FOB is made common knowledge *after* the signal θ_j has been transmitted, but *before* the dictator makes the choice.[25] As we now show, our model predicts that, if the dictator cares about the attribution of intentionality (i.e. $\lambda_2 > 0$), the manipulation of the recipient's knowledge will change his behavior.

Analogously as in the previous section, we denote the simple and intentional surprise functions given signal θ_j as $S_i^S(t_i, \theta_j)$ and $S_i^I(t_i, \theta_j)$, respectively, and the utility function as $U_i(t_i, \theta_j)$. Consider the difference in the simple surprise $S_i^S(t_i, \theta_j)$ between the treatments. Since the treatment manipulation occurs *after* the signal θ_j is transmitted, the *ex ante* recipient's FOB, signaled by θ_j, is equivalent in both treatments. Consequently, the simple surprise $S_i^S(t_i, \theta_j)$, which incorporates only the dictator's belief about the ex ante recipient's FOB, does not vary between the treatments (in what follows the lower index *pub* stays for the PUBLIC treatment, and *pr* for the PRIVATE treatment):

$$S_{i,pub}^S(t_i, \theta_j) = S_{i,pr}^S(t_i, \theta_j). \tag{11}$$

In contrast, the intentional surprise $S_i^I(t_i, \theta_j)$ is based on the *ex post* third- and fourth-order belief, and may thus depend on the interim treatment manipulation. Consider the intentional surprise in the PUBLIC treatment. Here the transmission of the signal θ_j is common knowledge, and hence the fourth-order belief of the dictator in the PUBLIC treatment is equal to his conditional SOB:

[25] Although this is typically not addressed in experimental economics, it is important to note that it is generally not possible to make sure that some information is actually common knowledge in the laboratory (because, e.g., somebody may have missed some information in the instructions). This is why we prefer the term *public* knowledge when we refer to experiment treatments, while we refer to the practically more demanding but theoretically simpler concept of common knowledge in our theory. In our analyses and in our experiment, the important aspect of the PUBLIC treatment is that a dictator knows that his recipient knows that he knows her first-order belief, which is what we explicitly told dictators in our laboratory if they participated in the PUBLIC treatment.

$$H_{ijij,pub}(x \mid t_i, \theta_j) = E_i E_j [H_{ij}(x \mid \theta_j)] = H_{ij}(x \mid \theta_j) \tag{12}$$

by the law of iterated conditional expectations (see Duffie (1988), p. 84). It follows, given (8) and (9), that in the PUBLIC treatment the intentional surprise $S^I_{i,pub}(t_i, \theta_j)$ is equal to the simple surprise $S^S_{i,pub}(t_i, \theta_j)$, that is, the recipient makes a correct inference about the intentions of the dictator:

$$S^I_{i,pub}(t_i, \theta_j) = S^S_{i,pub}(t_i, \theta_j).^{26} \tag{13}$$

Next, consider the intentional surprise in the PRIVATE treatment. In this case the dictator's SOB (formed by the observed guess θ_j) is not common knowledge as before. Therefore, the recipient's ex post third-order belief about the dictator's SOB can be different from the actual dictator's SOB.

We model the formation of the third-order belief in the PRIVATE treatment in the following way. Denote by $H^0_{ij}(\cdot)$ the cdf of the dictator's ex ante SOB (with pdf $h^0_{ij}(\cdot)$), which the dictator would have if he did not have any prior information about the recipient's FOB (that the recipient believes to be indeed so in the PRIVATE treatment). The recipient has uncertainty about $H^0_{ij}(\cdot)$ so that her (unconditional) third-order belief is represented by a probability weighting over possible dictator's ex ante SOBs:

$$H_{jij}(x) = \sum_K H^0_{ij,\kappa}(x)\, p_{j,\kappa}\,, \tag{14}$$

where $\kappa \in K$ is a parameter indexing the family of possible ex ante SOB distributions, and $p_{j,\kappa}$ is the unconditional probability of $H^0_{ij,\kappa}$ expected by the recipient. This uncertainty can be justified by heterogeneity of subjects' beliefs, for which empirical evidence was obtained, e.g., in our Experiment 1: the standard

[26] Given (13) and the fact that the dictator's knowledge about the recipient's transfer cannot (and actually was not) kept private ex post in a classroom experiment, the prediction of the generalized model in this section for Experiment 1 is the same as of the simpler model based on the simple surprise only (with utility given by (2)).

deviations of recipients' guesses and dictators' self-reported SOBs are 56% and 54% of the respective means. Hence, as before, we do not directly impose any consistency restrictions *between* the recipients' and the dictators' beliefs. However, we impose restrictions on the *internal* consistency of beliefs. In particular, we assume that the recipient believes that the ex ante SOB of the dictator is internally consistent, i.e. represents a consistent assessment (as in Battigalli and Dufwenberg (2009)). That is, the recipient believes that $H_{ij}^0(\cdot)$ should be (expectedly) unbiased relative to the distribution of transfers conditional on $H_{ij}^0(\cdot)$. Formally, this is expressed in the form of the following assumption:

A5. $H_j(x \mid H_{ij,\kappa}^0) = H_{ij,\kappa}^0(x)$ for any $x \in [0,1]$.

We also assume that the set of dictators' ex ante SOB distributions has a strict monotone likelihood ratio property (MLRP):

A6. $h_{ij}^0(\cdot)$ can be ordered by some indexing parameter γ so that

$$\frac{d}{dx}\left(\frac{h_{ij,\gamma_2}^0(x)}{h_{ij,\gamma_1}^0(x)}\right) > 0 \text{ if and only if } \gamma_2 > \gamma_1 \text{ for any } x \in [0,1].$$

This assumption roughly implies that for any two possible SOB distributions one distribution can be said to reflect higher beliefs than the other.

Regarding the pdf of the ex ante SOB, we assume:

A7. $h_{ij,\kappa}^0(x)$ is strictly positive and differentiable on $[0,1]$ for any κ.

Further, for simplicity, we consider the case where the dictator assigns probability 1 to the fact that the recipient considers only two possible SOB distributions with cdfs $H_{21}^0(\cdot)$ and $H_{22}^0(\cdot)$ (and pdfs $h_{21}^0(\cdot)$ and $h_{22}^0(\cdot)$)[27], which have ex ante probabilities p_1 and $p_2 = 1 - p_1$. The ex ante probabilities can be inferred by the dictator through observing θ_j (given our assumptions there exists a one-to-one correspondence from

[27] The first lower index stands for the second order of belief.

p_1 to θ_j, i.e. to the median of the ex ante FOB distribution). Assumption A6 for this case translates into

$$\left(\frac{h_{22}^0(x)}{h_{21}^0(x)}\right)' > 0 \tag{15}$$

for any $x \in [0,1]$, where the order of functions is without loss of generality.

Let us now consider the update of the recipient's third-order beliefs after observing the transfer. Denoting by $p_{1|t_i}$ the posterior probability that the recipient allocates to $h_{21}^0(\cdot)$ after observing transfer t_i, we have by Bayes rule

$$p_{1|t_i} = \frac{h_j(t_i \mid h_{21}^0)p_1}{h_j(t_i \mid h_{21}^0)p_1 + h_j(t_i \mid h_{22}^0)(1-p_1)} = \frac{h_{21}^0(t_i)p_1}{h_{21}^0(t_i)p_1 + h_{22}^0(t_i)(1-p_1)}, \tag{16}$$

where the last equality is due to Assumption A5. Since the dictator can infer p_1 from θ_j, he can thus infer $p_{1|t_i}$ as well. Thus, both the third-order belief of the recipient and the fourth-order belief of the dictator are given by

$$H_{ijij}(x \mid t_i) = H_{jij}(x \mid t_i) = H_{21}^0(x)p_{1|t_i} + H_{22}^0(x)(1-p_{1|t_i}). \tag{17}$$

We normalize the belief updating at the transfer equal to the median of the recipient's FOB distribution θ_j, by assuming that there is no update in the recipient's unconditional third-order belief in the case when the observed transfer corresponds to θ_j:[28]

A8. $H_{jij}(x \mid t_i = \theta_j) = H_{jij}(x)$.

[28] This assumption is justified given that in our experimental design θ_j will be estimated by the best recipient's guess of the average transfer. In this sense, the transfer equal to θ_j reconfirms the recipient's belief. In terms of our example with two densities, Assumption A8 is formally equivalent to $h_{21}^0(\theta_j) = h_{22}^0(\theta_j)$ (i.e. the densities intersect at θ_j), in which case the transfer equal to θ_j will not cause an update of the third-order belief.

Thus, in the PRIVATE treatment the fourth-order belief becomes endogenous to the transfer, since the latter signals the dictator's SOB (as believed by the recipient). This fact makes it different from the fourth-order belief in the PUBLIC treatment with common knowledge about the actual dictator's SOB, where the third- and fourth-order beliefs do not depend on the observed transfer. This implies that the treatment variation affects the intentional surprise $S_i^I(t_i, \theta_j)$, which in turn, as we show below, leads to a smaller transfer in the PRIVATE treatment.

Our argumentation (presented formally in Appendix B) is mainly based on the FOSD property of the posterior beliefs in the PRIVATE treatment:

Lemma 1. *In the PRIVATE treatment the posterior fourth-order belief* $H_{ijij}(x \mid t_i)$ *exhibits a strict FOSD in* t_i*, so that* $\frac{\partial H_{ijij}(x \mid t_i)}{\partial t_i} < 0$ *for any on* $x \in (0,1)$ *and* $t_i \in [0,1]$.

Proof. See Appendix B. ■

Intuitively, the recipient believes that in equilibrium transfers should be consistent with dictators' SOBs (in line with Assumption A5). Hence, after observing a relatively high transfer, the recipient is more likely to attribute the transfer to the dictator's higher SOB, and this is in turn anticipated by the dictator. Importantly, the positive relation between the recipient's belief and the transfer holds for both, relatively guilt averse as well as relatively surprise seeking dictators, and is in both cases driven by the assumption of the anticipated internal consistency of beliefs.

Lemma 1 implies that there is less scope for attribution of intentions in the PRIVATE treatment: the recipient does not know in PRIVATE whether a low (high) transfer is given with full awareness of the corresponding letting down (positive surprise), or whether it is just an artifact of the dictator's own low (high) SOB. Hence, the dictator's incentives to give positive transfers are reduced, resulting in a generally lower transfer in the PRIVATE treatment. This result is stated in the following

proposition (where $t^*_{i,pr}(\theta_j)$ and $t^*_{i,pub}(\theta_j)$ denote the optimal transfers conditional on signal θ_j in the PRIVATE and PUBLIC treatment, respectively):[29]

Proposition 3. *If* $0 < t^*_{i,pub}(\theta_j) < 1$, *then* $t^*_{i,pr}(\theta_j)$ *is strictly lower than* $t^*_{i,pub}(\theta_j)$.

Proof. See Appendix B. ■

Proposition 3 is robust. First, it holds independently of to what extent dictators care for negative and positive surprises, and, in particular, of the sign of $\alpha_i - \beta_i$.[30] Second, here too, behavior cannot be confounded by potential false consensus effects, because the dictator is informed of the recipient's FOB in both treatments. This leaves no room for the false consensus. Finally, because we know that false consensus may well matter, we note that we also do not expect the effect in Proposition 3 being weakened if players take a potential false consensus effect into account (see Appendix C). Thus, introducing the attribution of intentions into the analysis allows for a clear and robust testable prediction of the role of others' expectations and intentions at the between-subject level (in contrast to the within-subject level considered in Experiment 1).

3.2. Experiment design and hypotheses

Our second experiment was conducted with 254 participants in the Cologne Laboratory for Economic Research and the Frankfurt Laboratory of Experimental Economics, again with undergraduates from economics and business. As in EJTT's original experiment, our recipients were asked about their expectation regarding the average amount a dictator would send. Each dictator was then informed about the expectation of the recipient matched to him before a decision on how to split 10 Euro is made.

[29] As in the previous section, we assume that the dictator plays a pure strategy conditional on θ_j.

[30] Some of the empirical results in this section indicate the presence of both kinds of surprising behavior; see below.

In line with our model, we conducted two treatments with a between-subject design. In the PUBLIC treatment, recipients were told, after the expectation was elicited, that the matched dictator would get to know their estimate before choosing the transfer. In the PRIVATE treatment, recipients were not informed that their estimations were communicated to dictators. The respective procedure was known to dictators. The design allows us to change the scope for the recipients' inferences on dictator intentions, while at the same time minimizing strategic reporting of expectations and keeping recipients' expectations fixed. Instructions can be found in Appendix D.

Our null hypothesis is that dictators are indifferent to recipients' inferences regarding the underlying intentions. Then, there should be no difference in transfers between the treatments. In contrast, Proposition 3 predicts that transfers are higher in the PUBLIC treatment.[31]

3.3. Results

The average amount sent in PUBLIC is with 1.68 Euro almost 70% higher than the 1.01 Euro observed in the PRIVATE treatment. A two-sided MWU-test comparing the distributions of transfers in PUBLIC and PRIVATE reveals a significant difference between the two treatments ($p = 0.022$).[32] This confirms the prediction of Proposition 3. Also, in line with the reduced incentive to take into account the expectation of the recipient in the PRIVATE treatment, we find that the share of subjects who deviate negatively (positively) from the transmitted expectation

[31] There is a large literature on the role of 'intentionality in reciprocity', as, for example, surveyed by Sobel (2005) and Cooper and Kagel (forthcoming), which would also suggest that intentions may matter. Our study differs from this literature in that it focuses on the motives for altruistic gift giving rather than on reciprocal interaction, and in that it manipulates intentionality by varying the commonality of knowledge rather than by varying strategy sets, payoff functions, or the decision maker. Moreover, this literature does not consider intentions to *surprise* others.

[32] When we compute transfers as a percentage of the total amount to be divided (the cake size differed across the experiments), we find that the average amounts sent are somewhat higher in Experiment 1 than in the PUBLIC treatment of Experiment 2 (23.2% versus 16.8%). Part of the reason might be the larger social distance between dictators and recipients in the laboratory in Experiment 2, compared to the classroom Experiment 1. However, this difference is not quite statistically significant ($p = 0.073$, two-sided MWU-test). Yet, if we compare Experiment 1 and the PRIVATE treatment (10.1% of the endowment sent), the difference becomes highly significant ($p < 0.001$, two-sided MWU-test).

is significantly higher (lower) in this condition ($p = 0.016$ and $p = 0.020$, respectively, two-sided Pearson χ^2-tests). 71.7% (43 out of 60) of the dictators in PRIVATE transfer less than the recipient's expectation compared to 50.7% of the dictators in PUBLIC (34 out of 67). In particular, positive surprising cannot be neglected in this experiment as well, as the share of dictators who exceed the recipient's expectation is more than twice as high in the PUBLIC treatment: 28.4% of the dictators in the PUBLIC treatment (19 out of 67) transfer more than the recipient's guess whereas 11.7% of the dictators do so in PRIVATE (7 out of 60).

Due to the heterogeneity of preferences, we should – similar to Experiment 1 and the related studies on guilt aversion – observe only little correlation between recipients' beliefs and transfers on the aggregate in both treatments. Indeed, this is what we find: Pearson correlation coefficients are 0.149 ($p = 0.230$) in the PUBLIC treatment and 0.020 ($p = 0.881$) in the PRIVATE treatment. These results are corroborated by Tobit models with the amount sent by the dictator as the dependent variable (Table 2). In Model 1, we only include a dummy variable for the PUBLIC treatment. Its coefficient is positive and highly significant, corroborating our treatment effect. In Model 2, we additionally include the expectation of the matched recipient which turns out to be insignificant while the PUBLIC dummy is largely unaffected. Finally, the strong treatment effect and the lack of an impact of the matched recipient's belief remain unchanged when we additionally include variables capturing the demographic backgrounds of the subjects (see Model 3).[33]

[33] The only significant impact of the demographical background is found for subjects who previously knew the decision situation of the dictator game – the negative and significant coefficient indicates that these subjects decrease their transfers compared to subjects without previous experience. Nobody participated twice in our experiments.

Table 2.
Determinants of dictator transfers in Experiment 2.

No.	1	2	3
Dependent Variable	Transfer	Transfer	Transfer
Model	Tobit	Tobit	Tobit
PUBLIC	0.982**	0.998**	1.017**
	[0.400]	[0.399]	[0.395]
Recipient's expectation		0.121	0.138
		[0.144]	[0.141]
Age			-0.009
			[0.056]
Female			-0.113
			[0.401]
Business student			-0.152
			[0.414]
Dictator game known			-0.799**
			[0.399]
Constant	0.353	0.076	0.752
	[0.303]	[0.452]	[1.619]
Observations	127	127	127
Log-Likelihood	-218.4	-218.0	-215.8

Tobit models are calculated to account for the share of observations with zero transfers. Standard errors are given in brackets. ** denotes significance at the 5%-level. 'Age' is the participant's age in years. The dummy variable 'Female' takes the value of 1 if the participant is female. 'Business student' is a dummy variable equal to 1 if the participant is enrolled at the faculty of economics and business. Finally, 'Dictator game known' is a dummy variable indicating whether the subject knew the decision situation in advance.

4. CONCLUSION

We propose a model of 'surprising gifts' to investigate the role of others' expectations for giving in dictator games. The model assumes that people care not only about negative but also about positive surprises induced by their actions. We find evidence for the model in two experiments. While, similar to EJTT, we do not find a correlation between induced SOBs and actual transfers on a between-subject level, a within-subject level analysis in the first experiment shows that a large fraction of dictators reacts to recipients' expectations. In particular, many dictators behave consistently with BD's notion of guilt aversion. Yet, there is also a significant share of dictators behaving consistently with a preference for exceeding others'

expectations. The heterogeneity of belief-dependent preferences among subjects explains the lack of correlation between SOBs and transfers in the aggregate.

We then extend our model to integrate the notion that dictators may care for the recipients' inferences about the intention behind a transfer. The model predicts lower incentives to transfer for both, relatively guilt-averse and relatively surprise-seeking dictators, if the inference about the dictator's intentions becomes ambiguous. Our data from the second experiment confirm that gift giving is belief-dependent and, additionally, that it is at least partly driven by an attribution effect.

Overall, our data are consistent with the hypothesis that guilt aversion is a major motivation for giving in dictator games. At the same time, our analysis highlights that many subjects also like to exceed others' expectations, and that taking this motive into account, along with a motivation that subjects care about the attribution of their intentions, may strongly improve the predictive value of the model.

Of course, finding that belief-dependent preferences play an important role in social behavior is not claiming that they play the *only* role. Others found, for instance, that assuming a preference for fair outcomes (Fehr and Schmidt (1999) and Bolton and Ockenfels (2000)), or other social motives such as adherence to others' giving behavior (Shang and Croson (2009)), is useful to organize social behavior in a wide variety of social contexts (see Cooper and Kagel, forthcoming, for a review). Thus, for one, it would be interesting to see in future research how our results generalize to other social contexts. In particular, it may be interesting to study how positive surprises affect the *behavior* of the surprised, and whether decision-makers can strategically make use of surprising gifts to induce favorable actions by others.[34] Second, it may be worthwhile to investigate in more detail how a concern to please the opponent interacts with preferences to adhere to general norms of social behavior. In our experimental context one might argue, for instance, that the recipient's belief, if transmitted to the dictator, may be a valuable signal of what is the generally acceptable transfer. If dictators are also driven by a preference to conform to such a generally accepted social norm – as opposed to a desire to please one's own, particular recipient – dictators have even more reason to condition their social

[34] See the example in Geanakoplos et al. (1989) how positive surprises can induce cooperation in a dynamic prisoner's dilemma game.

behavior on beliefs. We note, however, that this view is difficult to reconcile with the negative correlation of induced SOB and transfers that we observe for many dictators in Experiment 1. Moreover, all that should matter for signaling the social norm is the knowledge about the recipient's expectation, while our Experiment 2 shows that, keeping everything else constant, dictators care about the recipient's attribution of their intentions to surprise, too. Clearly, beliefs about one's recipient's expectation and the attribution of intentions affect gift-giving behavior of a substantial share of dictators.

REFERENCES

Andreoni, J., and B. D. Bernheim (2009): "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects," *Econometrica*, 77, 1607–1636.

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

Battigalli, P., and M. Dufwenberg (2009): "Dynamic Psychological Games," *Journal of Economic Theory*, 141, 1–35.

Bellemare, C., A. Sebald, and M. Strobel (2011): "Measuring the Willingness to Pay to Avoid Guilt: Estimation Using Equilibrium and Stated Belief Models," *Journal of Applied Econometrics*, 26, 437–453.

Bénabou, R., and J. Tirole (2006): "Incentives and Prosocial Behavior," *American Economic Review*, 96, 1652–1678.

Bolton, G. E., and A. Ockenfels (2000): "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 90, 166–193.

Brandts, J., and G. Charness (2011): "The Strategy Versus the Direct-Response Method: A First Survey of Experimental Comparisons," *Experimental Economics*, 14, 375–398.

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

Charness, G., and M. Dufwenberg (2011): "Participation," *American Economic Review*, 101, 1211–1237.

Cooper, D.J., and J. H. Kagel, forthcoming: "Other-Regarding Preferences: A Selective Survey of Experimental Results," in *Handbook of Experimental Economics*, Vol. 2, ed. by J. H. Kagel and A. E. Roth. Princeton University Press.

Dana, J., R.A. Weber, and J. A. Kuang (2007): "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness," *Economic Theory*, 33, 67–80.

Dawes, R.M. (1989): "Statistical Criteria for Establishing a Truly False Consensus Effect," *Journal of Experimental Social Psychology*, 25, 1–17.

Duffie, D. (1988): *Security Markets: Stochastic Models*. Boston: Academic Press.

Dufwenberg, M., S. Gächter, and H. Hennig-Schmidt (2011): "The Framing of Games and the Psychology of Play," *Games and Economic Behavior*, 73, 459–478.

Edlin, A. S., and C. Shannon (1998): "Strict Monotonicity in Comparative Statics," *Journal of Economic Theory*, 81, 201–219.

Ellingsen, T., M. Johannesson, S. Tjøtta, and G. Torsvik (2010): "Testing Guilt Aversion," *Games and Economic Behavior*, 68, 95–107.

Falk, A., and M. Kosfeld (2006): "The Hidden Costs of Control," *The American Economic Review*, 96, 1611–1630.

Fehr, E., and K. M. Schmidt (1999): "A Theory of Fairness, Competition and Cooperation," *Quarterly Journal of Economics*, 114, 817–868.

Fischbacher, U., S. Gächter, and S. Quercia (2012): "The Behavioral Validity of the Strategy Method in Public Good Experiments," *Journal of Economic Psychology*, 33, 897–913.

Forsythe, R., J. Horowitz, N. E. Savin, and M. Sefton (1994): "Fairness in Simple Bargaining Experiments," *Games and Economic Behavior*, 6, 347–369.

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

Greene, W. (2004): "Fixed Effects and Bias Due to the Incidental Parameters Problem in the Tobit Model," *Econometric Reviews*, 23, 125–147.

Grossman, Z. (2010): "Self-Signaling Versus Social Signaling in Giving," University of California at Santa Barbara, Economics Working Paper Series.

Guerra, G., and D. J. Zizzo (2004): "Trust Responsiveness and Beliefs," *Journal of Economic Behavior and Organization*, 55, 25–30.

Kahneman, D., and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–291.

Köszegi, B., and M. Rabin (2006): "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 121, 1133–1165.

Mellers, B. A., A. Schwartz, K. Ho, and I. Ritov (1997): "Decision Affect Theory: Emotional Reactions to the Outcomes of Risky Options," *Psychological Science*, 8, 423–429.

Milgrom, P., and C. Shannon (1994): "Monotone comparative statics," *Econometrica*, 64, 157–181.

Ockenfels, A., and P. Werner (2012): "'Hiding Behind a Small Cake' in a Newspaper Dictator Game," *Journal of Economic Behavior and Organization*, 82, 82–85.

Ross, L., D. Greene, and P. House (1977): "The "False Consensus Effect": An Egoistic Bias in Social Perception and Attribution Processes," *Journal of Experimental Social Psychology*, 13, 279–301.

Selten, R. (1967): "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments," in *Beiträge zur experimentellen Wirtschaftsforschung*, ed. by H. Sauermann. Tübingen, Germany: J.C.B. Mohr (Paul Siebeck), 136–168.

Selten, R., and A. Ockenfels (1998): "An Experimental Solidarity Game," *Journal of Economic Behavior and Organization*, 34, 517–539.

Shang, J., and R. Croson (2009): "A Field Experiment in Charitable Contribution: The Impact of Social Information on the Voluntary Provision of Public Goods," *The Economic Journal*, 119, 1422–1439.

Sobel, J. (2005): "Interdependent Preferences and Reciprocity," *Journal of Economic Literature*, 43, 392–436.

Tversky, A., and D. Kahneman (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 297–323.

Vanberg, C. (2008): "Why do People Keep their Promises? An Experimental Test of Two Explanations," *Econometrica*, 76, 1467–1480.

APPENDIX A. Omitted proofs: Experiment 1.

Proof of Proposition 1. Given that $t_i^*(\theta_j) > 0$, the marginal utility should be (weakly) positive at $t_i^*(\theta_j)$.[35] Hence, we have (suppressing the argument in $t_i^*(\theta_j)$ for notational simplicity):

$$\begin{aligned}\frac{\partial U_i}{\partial t_i}(t_i^*, \theta_j) &= -m_i'(1-t_i^*) + \alpha_i \int_0^{t_i^*} h_{ij}(x \mid \theta_j)dx + \beta_i \int_{t_i^*}^1 h_{ij}(x \mid \theta_j)dx \\ &= (\alpha_i - \beta_i)H_{ij}(t_i^* \mid \theta_j) + \beta_i - m_i'(1-t_i^*) \geq 0.\end{aligned} \tag{18}$$

It follows from Assumption A2 that

$$(\alpha_i - \beta_i)H_{ij}(t_i^* \mid \theta_j) + \beta_i - \frac{\alpha_i + \beta_i}{2} \geq 0, \tag{19}$$

which is equivalent to

$$\frac{1}{2}(\alpha_i - \beta_i)\left(2H_{ij}(t_i^* \mid \theta_j) - 1\right) \geq 0. \tag{20}$$

Consequently, if $\alpha_i > \beta_i$, then $H_{ij}(t_i^* \mid \theta_j) \geq \frac{1}{2} = H_{ij}(\theta_j \mid \theta_j)$, where the last equality is due to θ_j coinciding with the median of $H_{ij}(\cdot \mid \theta_j)$ by (4). Thus, if $\alpha_i > \beta_i$, then $t_i^* \geq \theta_j$. By the analogous argument, if $\beta_i > \alpha_i$, then $t_i^* \leq \theta_j$.[36] ■

Lemma 2. *$U_i(t_i, \theta_j)$ has the strict single crossing property[37] in (t_i, θ_j) ($(t_i, -\theta_j)$) if $\alpha_i < \beta_i$ ($\alpha_i > \beta_i$).*

Proof. Taking the partial derivative of $\partial U_i / \partial t_i$ (see (18)) with respect to θ_j we get

[35] I.e., the marginal utility is 0 if $0 < t_i^*(\theta_j) < 1$ and is (weakly) larger than 0 if $t_i^*(\theta_j) = 1$.

[36] If $\alpha_i = \beta_i$, then (20) does not preclude the optimal transfer to be at any value.

[37] Function $f(x, z)$ has the strict single crossing property in (x, z) if for any $x' > x''$ and $z' > z''$ it holds that $f(x', z'') - f(x'', z'') \geq 0$ implies $f(x', z') - f(x'', z') > 0$ (Milgrom and Shannon (1994)).

$$\frac{\partial^2 U_i(t_i,\theta_j)}{\partial t_i \partial \theta_j} = (\alpha_i - \beta_i)\frac{\partial H_{ij}(t_i \mid \theta_j)}{\partial \theta_j}. \tag{21}$$

This, together with (5), implies that

$$\operatorname{sgn}\left(\frac{\partial^2 U_i(t_i,\theta_j)}{\partial t_i \partial \theta_j}\right) = -\operatorname{sgn}(\alpha_i - \beta_i) \tag{22}$$

for any $t_i \in (0,1)$, which leads to the claim. ■

Proof of Proposition 2. Let us consider arbitrary values θ_j' and θ_j'' such that $\theta_j' > \theta_j''$. By Lemma 2 and Milgrom-Shannon (1994) Monotone Selection Theorem we have that for any $t_i' \in \arg\max_{t_i} U_i(t_i,\theta_j')$ and $t_i'' \in \arg\max_{t_i} U_i(t_i,\theta_j'')$ it holds $t_i' \geq t_i''$ $(t_i' \leq t_i'')$ if $\alpha_i < \beta_i$ $(\alpha_i > \beta_i)$. That is, since $t_i^*(\theta_j)$ is a best response function such that $t_i^*(\theta_j) \in \arg\max_{t_i} U(t_i,\theta_j)$, it is weakly increasing (decreasing) in θ_j if $\alpha_i < \beta_i$ $(\alpha_i > \beta_i)$. If, in addition, $0 < t_i'' < 1$, then the FOC for maximizing $U_i(t_i,\theta_j)$ must be satisfied at (t_i'',θ_i''):

$$\frac{\partial U_i}{\partial t_i}(t_i'',\theta_i'') = 0. \tag{23}$$

Equation (23), together with (22), implies that if $\alpha_i \neq \beta_i$, then

$$\frac{\partial U_i}{\partial t_i}(t_i'',\theta_j') \neq 0. \tag{24}$$

It follows that $t_i'' \notin \arg\max_{t_i} U_i(t_i,\theta_j')$, hence $t_i'' \neq t_i'$ (if $\alpha_i \neq \beta_i$).[38] This, together with the previously established fact that $t_i' \leq t_i''$ $(t_i' \geq t_i'')$ if $\alpha_i < \beta_i$ $(\alpha_i > \beta_i)$, yields that $t_i^*(\theta_j)$ is *strictly* increasing (decreasing) in θ_j if $\alpha_i < \beta_i$ $(\alpha_i > \beta_i)$ and $0 < t_i^*(\theta_j) < 1$. ■

[38] The argument is analogous to the proof of Theorem 1 in Edlin and Shannon (1998).

APPENDIX B. Omitted proofs: Experiment 2.

Let us simplify the subsequent notation so that the pdf of belief of order k is denoted as $h_k(\cdot)$ and the cdf as $H_k(\cdot)$.[39]

Proof of Lemma 1. Substituting (16) into (17) we get

$$H_4(x \mid t_i) = \frac{p_1 H_{21}^0(x) h_{21}^0(t_i) + (1-p_1) H_{22}^0(x) h_{22}^0(t_i)}{p_1 h_{21}^0(t_i) + (1-p_1) h_{22}^0(t_i)}. \tag{25}$$

Taking the derivative of $H_4(x \mid t_i)$ with respect to t_i we obtain

$$\frac{\partial H_4(x \mid t_i)}{\partial t_i} = \frac{p_1(1-p_1)(H_{22}^0(x) - H_{21}^0(x))\left(\frac{dh_{22}^0(t_i)}{dt_i} h_{21}^0(t_i) - \frac{dh_{21}^0(t_i)}{dt_i} h_{22}^0(t_i) \right)}{(p_1 h_{21}^0(t_i) + (1-p_1) h_{22}^0(t_i))^2}. \tag{26}$$

Note that by Assumption A7 this derivative always exists. Consider the nominator of (26). We have $H_{22}^0(x) - H_{21}^0(x) < 0$ for $x \in (0,1)$, while the strict MLRP (15) implies strict FOSD of H_{22}^0 over H_{21}^0. Besides, (15) yields $\frac{dh_{22}^0(t_i)}{dt_i} h_{21}^0(t_i) - \frac{dh_{21}^0(t_i)}{dt_i} h_{22}^0(t_i) > 0$ for any $t_i \in [0,1]$. Consequently,

$$\frac{\partial H_4(x \mid t_i)}{\partial t_i} < 0 \tag{27}$$

for any $x \in (0,1)$ and $t_i \in [0,1]$. ■

Lemma 3. $\frac{\partial S_{i,pr}^I(t_i, \theta_j)}{\partial t_i} \le \frac{\partial S_{i,pub}^I(t_i, \theta_j)}{\partial t_i}$ *if* $t_i \le \theta_j$ *and* $\beta_i \ge \alpha_i$, *or if* $t_i \ge \theta_j$ *and* $\alpha_i \ge \beta_i$, *with a strict inequality if, in addition,* $0 < t_i < 1$.

Proof. To avoid notational confusion, let us prove the claim of the proposition for a given value of transfer $t_i = \tilde{t}$ with $0 \le \tilde{t} \le 1$.

[39] The previous notation for $h_{2k}^0(\cdot)$ and $H_{2k}^0(\cdot)$, $k = 1,2$, is left unchanged.

Let us first consider the intentional surprise in the PUBLIC treatment $S^I_{i,pub}(\tilde{t},\theta_j)$. Given (13) and (8) we have

$$\begin{aligned} S^I_{i,pub}(\tilde{t},\theta_j) &= S^S_i(\tilde{t},\theta_j) = \alpha_i\int_0^{\tilde{t}}(\tilde{t}-x)h_2(x\,|\,\theta_j)dx - \beta_i\int_{\tilde{t}}^1(x-\tilde{t})h_2(x\,|\,\theta_j)dx \\ &= \alpha_i\int_0^{\tilde{t}}H_2(x\,|\,\theta_j)dx + \beta_i\int_{\tilde{t}}^1 H_2(x\,|\,\theta_j)dx - \beta_i\left(1-\tilde{t}\right), \end{aligned} \tag{28}$$

where the last line is obtained by integration by parts as in (21). Taking the derivative we get

$$\frac{\partial S^I_{i,pub}(\tilde{t},\theta_j)}{\partial\tilde{t}} = (\alpha_i-\beta_i)H_2(\tilde{t}\,|\,\theta_j)+\beta_i. \tag{29}$$

Next, note that the unconditional third-order belief of the recipient is consistent with her unconditional FOB:

$$H_3(x) = \sum_K H^0_{2,\kappa}(x)\,p_{j,\kappa} = \sum_K H_1(x\,|\,H^0_{2,\kappa})\,p_{j,\kappa} = H_1(x), \tag{30}$$

where the first equality is by (14), the second by Assumption A5 and the third by the law of total probability. Then,

$$\begin{aligned} H_2(x\,|\,\theta_j) &= E_i[H_1(x)\,|\,\theta_j] = E_i[H_3(x)\,|\,\theta_j] \\ &= E_i[H_3(x\,|\,t_i=\theta_j)] = H_4(x\,|\,t_i=\theta_j), \end{aligned} \tag{31}$$

where the second equality is by (30) and the third by Assumption A8. This, together with (29), yields

$$\frac{\partial S^I_{i,pub}(\tilde{t},\theta_j)}{\partial\tilde{t}} = (\alpha_i-\beta_i)H_4(\tilde{t}\,|\,t_i=\theta_j)+\beta_i. \tag{32}$$

At the same time, applying integration by parts[40], we have the following expression for the intentional surprise in the PRIVATE treatment $S^I_{i,pr}(\tilde{t},\theta_j)$ (generally given by (9)):

[40] The fact that $H_4(x\,|\,t_i)$ is continuously differentiable in x follows from Assumption A7.

$$S^I_{i,pr}(\tilde{t},\theta_j) = \alpha_i \int_0^{\tilde{t}} (\tilde{t}-x) h_4(x\,|\,\tilde{t})dx - \beta_i \int_{\tilde{t}}^1 (x-\tilde{t}) h_4(x\,|\,\tilde{t})dx$$
$$= \alpha_i \int_0^{\tilde{t}} H_4(x\,|\,\tilde{t})dx + \beta_i \int_{\tilde{t}}^1 H_4(x\,|\,\tilde{t})dx - \beta_i (1-\tilde{t}). \quad (33)$$

Taking the derivative yields

$$\frac{\partial S^I_{i,pr}(\tilde{t},\theta_j)}{\partial \tilde{t}} = (\alpha_i - \beta_i) H_4(\tilde{t}\,|\,\tilde{t}) + \beta_i$$
$$+\alpha_i \int_0^{\tilde{t}} \frac{\partial H_4(x\,|\,\tilde{t})}{\partial \tilde{t}} dx + \beta_i \int_{\tilde{t}}^1 \frac{\partial H_4(x\,|\,\tilde{t})}{\partial \tilde{t}} dx. \quad (34)$$

Subtracting (32) from (34) we arrive at

$$\frac{\partial S^I_{i,pr}(\tilde{t},\theta_j)}{\partial \tilde{t}} - \frac{\partial S^I_{i,pub}(\tilde{t},\theta_j)}{\partial \tilde{t}} = D_1 + D_2, \quad (35)$$

where

$$D_1 = (\alpha_i - \beta_i)(H_4(\tilde{t}\,|\,\tilde{t}) - H_4(\tilde{t}\,|\,t_i = \theta_j)), \quad (36)$$

$$D_2 = \alpha_i \int_0^{\tilde{t}} \frac{\partial H_4(x\,|\,\tilde{t})}{\partial \tilde{t}} dx + \beta_i \int_{\tilde{t}}^1 \frac{\partial H_4(x\,|\,\tilde{t})}{\partial \tilde{t}} dx. \quad (37)$$

We have $D_1 \le 0$ by Lemma 1 and initial conditions. At the same time, Lemma 1 also implies that $D_2 \le 0$, with a strict inequality if $0 < \tilde{t} < 1$ (given Assumption A1). Consequently, the LHS of (35) is weakly negative, being strictly negative if $0 < \tilde{t} < 1$. ■

Corollary 1. $\frac{\partial U_{i,pr}(t_i,\theta_j)}{\partial t_i} \le \frac{\partial U_{i,pub}(t_i,\theta_j)}{\partial t_i}$ *if* $t_i \le \theta_j$ *and* $\beta_i \ge \alpha_i$, *or if* $t_i \ge \theta_j$ *and* $\alpha_i \ge \beta_i$, *with a strict inequality if, in addition,* $0 < t_i < 1$.

Proof. Given that the difference between the treatments affects only $S^I_i(t_i,\theta_j)$ (see (11)), the claim follows directly by Lemma 3. ■

Corollary 2. *For any* t' *and* t'' *so that* $t'' < t' \le \theta_j$ *and* $\beta_i \ge \alpha_i$, *or* $\theta_j \le t'' < t'$ *and* $\alpha_i \ge \beta_i$, *it holds:*

$$U_{i,pub}(t',\theta_j) - U_{i,pub}(t'',\theta_j) > U_{i,pr}(t',\theta_j) - U_{i,pr}(t'',\theta_j).$$

Proof. The claim follows from the fact that $U_i(t',\theta_j) - U_i(t'',\theta_j) = \int_{t''}^{t'} \frac{\partial U_i(y,\theta_j)}{\partial y} dy$ and Corollary 1. ■

Lemma 4. $U_{i,pr}(t_i,\theta_j) > U_{i,pub}(t_i,\theta_j)$ *if* $0 < t_i < \theta_j$, *and* $U_{i,pr}(t_i,\theta_j) \le U_{i,pub}(t_i,\theta_j)$ *if* $t_i \ge \theta_j$.

Proof. We have by (11)

$$U_{i,pr}(t_i,\theta_j) - U_{i,pub}(t_i,\theta_j) = \lambda_2 (S^I_{i,pr}(t_i,\theta_j) - S^I_{i,pub}(t_i,\theta_j)). \tag{38}$$

At the same time, it follows from (28) and (31) that, denoting $H_4(x \mid t_i = \theta_j)$ as $H_4(x \mid \theta_j)$,

$$S^I_{i,pub}(t_i,\theta_j) = \alpha_i \int_0^{t_i} H_4(x \mid \theta_j) dx + \beta_i \int_{t_i}^1 H_4(x \mid \theta_j) dx - \beta_i (1 - t_i). \tag{39}$$

Subtracting (39) from (33) yields

$$\begin{aligned} & S^I_{i,pr}(t_i,\theta_j) - S^I_{i,pub}(t_i,\theta_j) \\ & = \alpha_i \int_0^{t_i} \left(H_4(x \mid t_i) - H_4(x \mid \theta_j) \right) dx + \beta_i \int_{t_i}^1 \left(H_4(x \mid t_i) - H_4(x \mid \theta_j) \right) dx. \end{aligned} \tag{40}$$

It follows from (40), Lemma 1 and Assumption A1 that $S^I_{i,pr}(t_i,\theta_j) - S^I_{i,pub}(t_i,\theta_j) > 0$ if $0 < t_i < \theta_j$, and $S^I_{i,pr}(t_i,\theta_j) - S^I_{i,pub}(t_i,\theta_j) \le 0$ if $t_i \ge \theta_j$. This, together with (38), leads to the claim. ■

Lemma 5. *If* $0 < t^*_{i,pub}(\theta_j) \le \theta_j$ *and* $\beta_i \ge \alpha_i$, *or if* $\theta_j \le t^*_{i,pub}(\theta_j) < 1$ *and* $\alpha_i \ge \beta_i$, *then* $t^*_{i,pr}(\theta_j) < t^*_{i,pub}(\theta_j)$.

Proof. For notational simplicity, let us suppress argument θ_j in the functions of optimal transfers $t^*_{i,pr}(\theta_j)$ and $t^*_{i,pub}(\theta_j)$. Let us first show the weak inequality $t^*_{i,pr} \le t^*_{i,pub}$ under the assumed conditions. Suppose to the contrary that $t^*_{i,pub} < t^*_{i,pr}$. Then, given the initial conditions there can exist only the following cases:

Case 1: $0 < t^*_{i,pub} < t^*_{i,pr} \le \theta_j$ and $\beta_i \ge \alpha_i$, or $\theta_j \le t^*_{i,pub} < t^*_{i,pr}$ and $\alpha_i \ge \beta_i$.

Case 2: $0 < t^*_{i,pub} < \theta_j < t^*_{i,pr}$ and $\beta_i \ge \alpha_i$.

Let us prove that both cases are contradictory.

Case 1:

By Corollary 2 it then follows

$$U_{i,pub}(t^*_{i,pr},\theta_j) - U_{i,pub}(t^*_{i,pub},\theta_j) > U_{i,pr}(t^*_{i.pr},\theta_j) - U_{i,pr}(t^*_{i.pub},\theta_j). \quad (41)$$

At the same time, given that $t^*_{i,pr}$ and $t^*_{i,pub}$ are the optimal choices in the respective treatments, we have

$$U_{i,pr}(t^*_{i,pr},\theta_j) - U_{i,pr}(t^*_{i,pub},\theta_j) \ge 0, \quad (42)$$

$$U_{i,pub}(t^*_{i,pr},\theta_j) - U_{i,pub}(t^*_{i,pub},\theta_j) \le 0. \quad (43)$$

This contradicts (41).

Case 2:

In this case, by Lemma 4

$$U_{i,pr}(t^*_{i,pub},\theta_j) > U_{i,pub}(t^*_{i,pub},\theta_j), \quad (44)$$

$$U_{i,pr}(t^*_{i,pr},\theta_j) \le U_{i,pub}(t^*_{i,pr},\theta_j). \quad (45)$$

At the same time, since $t^*_{i,pub}$ is the optimal transfer in the PUBLIC treatment, it holds

$$U_{i,pub}(t^*_{i,pr},\theta_j) \le U_{i,pub}(t^*_{i,pub},\theta_j). \quad (46)$$

It follows from (44), (45) and (46) that

$$U_{i,pr}(t^*_{i,pr},\theta_j) < U_{i,pr}(t^*_{i,pub},\theta_j), \quad (47)$$

contradicting to $t^*_{i,pr}$ being the optimal transfer in the PRIVATE treatment.

Thus, we have come to contradiction in all possible cases when $t^*_{i,pub} < t^*_{i,pr}$, hence

$$t^*_{i,pr} \le t^*_{i,pub}. \quad (48)$$

Moreover, this inequality is strict since $0 < t^*_{i,pub} < 1$ by assumption. Indeed, in this case FOC for $t^*_{i,pub}$ is satisfied, i.e.

$$\frac{\partial U_{i,pub}}{\partial t_i}(t^*_{i,pub},\theta_j) = 0. \quad (49)$$

By Corollary 1 it then follows

$$\frac{\partial U_{i,pr}}{\partial t_i}(t^*_{i,pub},\theta_j) < 0. \quad (50)$$

This means that $U_{i,pr}(t_i,\theta_j)$ is strictly decreasing at $t_i = t^*_{i,pub}$, implying together with (48) that $t^*_{i,pr} < t^*_{i,pub}$. ■

Proof of Proposition 3. Given that $0 < t^*_{i,pub}(\theta_j) < 1$, the following first-order condition for maximizing utility in the PUBLIC treatment must be satisfied (suppressing the argument in $t^*_i(\theta_j)$ for notational simplicity):

$$\begin{aligned}
&\frac{\partial U_{i,pub}}{\partial t_i}(t^*_{i,pub},\theta_j) = -m_i'(1-t^*_{i,pub}) \\
&+(\lambda_1+\lambda_2)\left(\alpha_i\int_0^{t^*_{i,pub}} h_2(x\,|\,\theta_j)dx + \beta_i\int_{t^*_{i,pub}}^1 h_2(x\,|\,\theta_j)dx\right) \\
&= -m_i'(1-t^*_{i,pub}) + \alpha_i\int_0^{t^*_{i,pub}} h_2(x\,|\,\theta_j)dx + \beta_i\int_{t^*_{i,pub}}^1 h_2(x\,|\,\theta_j)dx \\
&= (\alpha_i-\beta_i)H_2(t^*_{i,pub}\,|\,\theta_j) + \beta_i - m_i'(1-t^*_{i,pub}) = 0,
\end{aligned} \tag{51}$$

where the first equality is by (13) and the second by (10). It follows from the last equality and Assumption A2 that

$$\frac{1}{2}(\alpha_i-\beta_i)\left(2H_2(t^*_{i,pub}\,|\,\theta_j)-1\right)\geq 0. \tag{52}$$

Consequently, given that θ_j is the median of $H_2(\cdot\,|\,\theta_j)$ by (4), if $t^*_{i,pub}>\theta_j$, then $\alpha_i\geq\beta_i$, and if $t^*_{i,pub}<\theta_j$, then $\beta_i\geq\alpha_i$. In these cases, $t^*_{i,pr}<t^*_{i,pub}$ by Lemma 5. At the same time, if $t^*_{i,pub}=\theta_j$, then the necessary condition for Lemma 5 is satisfied for any α_i and β_i. Consequently, $t^*_{i,pr}<t^*_{i,pub}$ as well. ■

APPENDIX C. Robustness of Proposition 3 to false consensus

One can show that the analytical results obtained in Subsection 3.1 are robust to the presence of false consensus, for which we found evidence in Experiment 1. In particular, we show below that if dictators care about the recipients' beliefs and, at the same time, do have a false consensus bias in their SOBs formation, then the same prediction for the difference in transfers between the treatments is obtained as under the consistent formation of beliefs (i.e., as under Assumption A5, which we lift here).

The false consensus typically implies that own behavior is considered as representative for the whole population (Ross et al. (1977)). In terms of our model, this means that one's own transfer serves as a signal about others' transfers and, more importantly, about others' representative expectations. This can be expressed in the

form of the assumption that, if no information about the recipient's belief is available, the dictator's SOB is formed in the same way as if he got a direct signal θ_j equal to his optimal transfer t_i^*:

$$\tilde{H}_{ij}(x \mid t_i^*) = H_{ij}(x \mid \theta_j = t_i^*), \tag{53}$$

where $\tilde{H}_{ij}(\cdot \mid t_i^*)$ is the cdf of the SOB under the false consensus. Besides, we assume that the recipient is aware of this fact, i.e. in the PRIVATE treatment her third-order belief is

$$H_{jij}(x \mid t_i) = E_j[\tilde{H}_{ij}(x \mid t_i^* = t_i)] = E_j[H_{ij}(x \mid \theta_j = t_i)]. \tag{54}$$

This awareness can be justified by the false consensus arising not from some irrational unconscious bias, but rather from a lack of dictator's information, whereby each dictator treats his transfer as the only signal to predict the behavior of others. Such kind of 'rational' false consensus was introduced by Dawes (1989).

Finally, we assume that the dictator is also aware of the recipient's updating process, i.e.

$$H_{ijij}(x \mid t_i) = E_i[H_{jij}(x \mid t_i)] = E_i E_j[H_{ij}(x \mid \theta_j = t_i)] = H_{ij}(x \mid \theta_j = t_i), \tag{55}$$

with the second equality by (54) and the third by the law of iterated conditional expectations. This implies

$$\frac{\partial H_{ijij}(x \mid t_i)}{\partial t_i} = \frac{\partial H_{ij}(x \mid \theta_j)}{\partial \theta_j}, \tag{56}$$

so that by Assumption A3 we have that $H_{ijij}(x \mid t_i)$ exhibits a strict FOSD in t_i:

$$\frac{\partial H_{ijij}(x \mid t_i)}{\partial t_i} < 0 \tag{57}$$

for any $x \in (0,1)$ and $t_i \in [0,1]$. Hence, we get the same result as in Lemma 1 for the previous case of rational beliefs. The main difference is that before the positive

correlation between transfers and beliefs was driven by the internal consistency of ex ante beliefs assumed by recipients (Assumption A5), while in the false consensus case beliefs are shaped by transfers directly.

The FOSD property of the fourth-order belief allows for the same line of reasoning by comparing the treatments as in the rational case. In the PRIVATE treatment the recipient updates her third-order belief so that it follows the observed transfer, since she believes that the dictator is uninformed and, thus, his SOB is subject to the false consensus effect. In contrast, the (presumed) false consensus effect does not affect beliefs in the PUBLIC treatment, since there we have common knowledge about the actual dictator's SOB (formed by signal θ_j). This leads to a smaller scope of attribution of intentions in the PRIVATE treatment, decreasing the dictator's corresponding motivations and yielding a smaller transfer, relatively to the level in the PUBLIC treatment. The formal proofs in this case follow the same lines as in Appendix B.[41]

We conclude that our analytical predictions from Subsection 3.1 and our hypothesis from Subsection 3.2 are robust to the presence of false consensus. Note finally that the false consensus effect alone clearly cannot cause any difference between the dictator's behavior in the PRIVATE and the PUBLIC treatment if the dictator does not care about the recipient's beliefs, i.e. if $\alpha_i = \beta_i = 0$ (the dictator would then simply follow the maximization of his material utility in both treatments).

[41] In particular, the result of Lemma 1 is directly stated by (57). The subsequent proof follows the same line of arguments as the proofs in Appendix B starting with Lemma 3 (with the only exception that (31) follows from (55) directly).

APPENDIX D. Experimental instructions

D.1 Experiment 1

Below you find instructions for Experiment 1 translated from German.

Dictators' instructions

General information

Welcome to our experiment. In this experiment you can earn money. You will receive your payoff against the attached receipt, which we ask you to keep.

You are not allowed to speak with other participants during the session. If you have any questions please raise your hand, the experimenter will come to help you. If you violate these rules, we will have to exclude you from the experiment and all payments.

Decision situation

In this experiment all participants are randomly divided into Participants A and Participants B, and each participant is randomly matched with another person. **You are Participant A, the other person is Participant B.**

Each pair receives an endowment of 14 Euro. Then you have to decide about how this sum should be divided between you and Participant B. This means, you determine your own amount and the amount of Participant B so that

Your payoff=14 Euro – amount of Participant B

Payoff of Participant B= amount of Participant B

Prior to your decision your matched Participant B will be asked to guess how much of the 14 Euro, on average, participant A will send to participant B. You will be informed about the guess of your matched Participant B first after your decision about the division of the sum is made. However, you can set the amount of Participant B to depend on the possible guesses of Participant B. The payoff-relevant amount of Participant B is then the amount that you chose for the actual guess of Participant B.

Participant B does not know that you will be informed about his guess and that you can condition your decision on it.

Participant B can get an additional payoff by his guess. The participant B, whose guess is the closest to the actual average amount received by Participants B, wins an additional bonus of 8.00 Euro. If several participants are closest, then the person who gets the bonus is determined randomly.
Take your time and make sure you understand these instructions.
All decisions and payoffs are confidential. No other participant will get to know your payoffs.
Moreover, no participant will get to know during or after the experiment which other participant he or she was assigned to.

Form for Participant A

Please indicate your decision.

If the (rounded) guess of my Participant B about his amount is the following (see inputs in this column)…	… then I give him the following amount:
0,00 €	_ _ , _ _ €
0,50 €	_ _ , _ _ €
1,00 €	_ _ , _ _ €
1,50 €	_ _ , _ _ €
2,00 €	_ _ , _ _ €
2,50 €	_ _ , _ _ €
3,00 €	_ _ , _ _ €
3,50 €	_ _ , _ _ €
4,00 €	_ _ , _ _ €
4,50 €	_ _ , _ _ €
5,00 €	_ _ , _ _ €
5,50 €	_ _ , _ _ €
6,00 €	_ _ , _ _ €
6,50 €	_ _ , _ _ €
7,00 €	_ _ , _ _ €
7,50 €	_ _ , _ _ €
8,00 €	_ _ , _ _ €
8,50 €	_ _ , _ _ €
9,00 € and more	_ _ , _ _ €

Post-experimental questionnaire

Finally, we would like to ask you a few questions.

Age: ____ years

Gender: (female/male)

Field of study: _____

Semester: _____

Mother tongue: _____

Do you know the decision situation from a previous experiment? (Yes/No)

What do you think is the amount that participant A should send to participant B? [0.00-14.00 Euro] _ _ , _ _ Euro.

What do you think is the average amount that participant A sends to participant B? [0.00-14.00 Euro] _ _ , _ _ Euro.

What do you think is the average guess of all participants B about the amount sent by participant A to participant B? [0.00-14.00 Euro] _ _ , _ _ Euro.

Recipients' instructions

General information

Welcome to our experiment. In this experiment you can earn money. You will receive your payoff against the attached receipt, which we ask you to keep.

You are not allowed to speak with other participants during the session. If you have any questions please raise your hand, the experimenter will come to help you. If you violate these rules, we will have to exclude you from the experiment and all payments.

Decision situation

In this experiment all participants are randomly divided into Participants A and Participants B, and each participant is randomly matched with another person. **You are Participant B, the other person is Participant A.**

Each pair receives an endowment of 14 Euro. Then Participant A has to decide about how this sum should be divided between himself and Participant B. This means, he determines his own amount and your amount so that

Payoff of Participant A=14 Euro – your amount
Your payoff = your amount

Before Participant A makes the decision, you will be asked to guess how much of the 14 Euro, on average, participant A will send to participant B.
You can get an additional payoff by your guess. The Participant B, whose guess is the closest to the actual average amount received by Participants B, wins an additional bonus of 8.00 Euro. If several participants are closest, then the person who gets the bonus is determined randomly.
Take your time and make sure you understand these instructions.
All decisions and payoffs are confidential. No other participant will get to know your payoffs.
Moreover, no participant will get to know during or after the experiment which other participant he or she was assigned to.

Form for Participant B

What do you believe is the average amount that Participants B will get?
Please state a value from [0.00 - 14.00 Euro]:

___ ___ , ___ ___ Euro.

Post-experimental questionnaire

Finally, we would like to ask you a few questions.
Age: ____ years
Gender: (female/male)
Field of study: _____

Semester: _____

Mother tongue: _____

Do you know the decision situation from a previous experiment? (Yes/No)

What do you think is the amount that participant A should send to participant B? [0.00-14.00 Euro] _ _ , _ _ Euro.

What is the amount that you would have sent in the role of participant A? [0.00-14.00 Euro] _ _ , _ _ Euro.

What do you think? What does your matched participant A think is the amount that participants B expect on average? [0.00-14.00 Euro] _ _ , _ _ Euro.

D.2 Experiment 2

Below you find instructions for Experiment 2 translated from German. The instructions for treatments PUBLIC and PRIVATE differ only in the sentences marked in the text.

Dictators' instructions

General information

Welcome to the experiment! In this experiment you can earn money. How much you can earn depends on your decisions. You will receive an amount of 2.50 Euro for your participation that will be paid out irrespective of the decisions in the experiment.

From now on please do not communicate with other participants. If you have a question, please raise your hand! We will come to your desk and answer your question. If you violate these rules, we will have to exclude you from the experiment and all payments.

Decision situation

In this experiment, two participants are randomly matched. One participant is randomly assigned the role of participant A, the other is randomly assigned the role of participant B.

You are participant A, the other person is participant B.

You receive an endowment of 10 Euro. From this endowment, you can send any amount to participant B. Payoffs are calculated as follows:

Your payoff = 10 Euro – amount sent

Payoff of B = Amount sent

Prior to your decision, participant B will be asked to guess how much of the 10 Euro, on average, participant A will send to participant B. You will be informed about participant B's guess before you decide on the amount to be sent.

[Treatment PUBLIC] After participant B made the guess he/she will be informed that you know his/her guess before you choose the amount to be sent.

[Treatment PRIVATE] Participant B will not be informed that you know his/her guess.

Participant B can achieve an additional payoff through his/her guess. The participant B whose guess is closest to the actual average will win an amount of 8 Euro.

Take your time and make sure that you understand these instructions. All decisions and payoffs are confidential. No participant will get to know your payoffs, and you will receive the money in a closed envelope when you leave the laboratory.

Moreover, no participant will get to know during or after the experiment which other participant he or she was assigned to.

Post-experimental questionnaire

Finally, we would like to ask you a few questions.

Age: ____ years

Gender: (female/male)

Faculty: (business/economics, law, medicine, arts and humanities, mathematics and natural sciences, human sciences, no student)

Semester: _____

Mother tongue: _____

Do you know the decision situation from a previous experiment? (Yes/No)

What do you think is the amount that participant A should send to participant B? [0.00-14.00 Euro] _ _ , _ _ Euro.
What do you think is the average amount that participant A sends to participant B? [0.00-14.00 Euro] _ _ , _ _ Euro.
What do you think is the average guess of all participants B about the amount sent by participant A? [0.00-14.00 Euro] _ _ , _ _ Euro.

Recipients' instructions

Instructions for player B were identical in the PUBLIC and the PRIVATE treatments. In the PUBLIC treatment, players B additionally received the following information displayed on the computer screens after they had typed in their guesses: "Participant A will be informed about your guess of ____Euro before he decides on the amount sent to you."

General information
Welcome to the experiment! In this experiment you can earn money. How much you can earn depends on your decisions. You will receive an amount of 2.50 Euro for your participation that will be paid out irrespective of the decisions in the experiment.
From now on please do not communicate with other participants. If you have a question, please raise your hand! We will come to your desk and answer your question. If you violate these rules, we will have to exclude you from the experiment and all payments.

Decision situation
In this experiment, two participants are randomly matched. One participant is randomly assigned the role of participant A, the other is randomly assigned the role of participant B.
You are participant B, the other person is participant A.
Participant A receives an endowment of 10 Euro. From this endowment, he or she can send any amount to you. Payoffs are calculated as follows:

Payoff of A = 10 Euro – amount sent

Your Payoff = Amount sent

Prior to participant A's decision, you will be asked to guess how much of the 10 Euro, on average, participant A will send to participant B.

You can achieve an additional payoff through your guess. The participant B whose guess is closest to the actual average will win an amount of 8 Euro.

Take your time and make sure that you understand these instructions. All decisions and payoffs are confidential. No participant will get to know your payoffs, and you will receive the money in a closed envelope when you leave the laboratory.

Moreover, no participant will get to know during or after the experiment which other participant he or she was assigned to.

Post-experimental questionnaire

Finally, we would like to ask you a few questions.

Age: ____ years

Gender: (female/male)

Faculty: (business/economics, law, medicine, arts and humanities, mathematics and natural sciences, human sciences, no student)

Semester: _____

Mother tongue: _____

Do you know the decision situation from a previous experiment? (Yes/No)

What do you think is the amount that participant A should send to participant B? [0.00-14.00 Euro] _ _ , _ _ Euro.

What is the amount that you would have sent in the role of participant A? [0.00-14.00 Euro] _ _ , _ _ Euro.

What do you think? What does your matched participant A think is the amount that participants B expect on average? [0.00-14.00 Euro] _ _ , _ _ Euro.