Descriptive Norms and Guilt Aversion*

Anastasia Danilov[†,a], Kiryl Khalmetski[‡,b], and Dirk Sliwka[§,b]

[a] *Humboldt-Universität zu Berlin, School of Business and Economics, Unter den Linden 6, 10099 Berlin, Germany*
[b] *University of Cologne, Department of Economics, Albertus-Magnus-Platz, 50923 Cologne, Germany*

October 13, 2021

Abstract

It has been argued that guilt aversion (the desire to meet others' expectations) and the social norm compliance (the desire to act similarly to other individuals in the same situation) are important drivers of human behavior. However, as we show in a theoretical model, these two motives are empirically indistinguishable when only one signal (either the expectation of a person affected by the choice or a signal about the descriptive norm) is revealed as each of these signals transmit information on the other benchmark. We address this problem by running an experiment in which signals for both benchmarks are revealed simultaneously. We find that both types of information affect dictator transfers in a one-shot game, yet the information about the behavior of others has a stronger effect than the disclosed recipient's expectation. The effect of the recipient's expectation is non-monotonic and becomes negative for very high expectations. We provide further evidence for the importance of guilt aversion in a second experiment where we display the recipient's expectation and the expectation of a randomly picked recipient of another dictator.

Keywords: guilt aversion, social norms, conformity, dictator game, experiment

JEL classification: C91, D83, D84

*This paper supersedes a previous paper titled "Norms and Guilt". We thank Loukas Balafoutas, Pierpaolo Battigalli, Gary Charness, Eugen Dimant, Martin Dufwenberg, Erin Krupka, Elena Manzoni, Roberto Weber and workshop participants in Soleto and Nice for helpful comments. Financial support of the German Research Foundation (DFG) through the Research Unit "Design and Behavior" (FOR 1371) and CRC TRR 190 "Rationality and competition"(Project number 280092119) is gratefully acknowledged.

†E-mail: anastasia.danilov@hu-berlin.de. *Corresponding author.*

‡E-mail: kk917@bath.ac.uk. *Present address*: University of Bath, Department of Economics, Claverton Down, Bath BA2 7JP, United Kingdom.

§E-mail: sliwka@wiso.uni-koeln.de.

1 Introduction

In situations involving social interactions, people tend to search for cues indicating suitable conduct in a given context. One such signal is the information about a typical behavior of other social group members in the same situation, or the *descriptive social norm* (Cialdini et al., 1990). Another important signal is the expectation of individuals affected by the choice (e.g., a waiter expecting a certain tip, an employee expecting a specific wage increase). A preference for compliance to such individual expectations is termed as *guilt aversion* (Battigalli and Dufwenberg, 2007).[1]

A large body of literature has acknowledged both of these mechanisms to be important drivers of individual behavior.[2] At the same time, there is still a significant gap in the understanding of how information about descriptive social norms and the expectation of others interact, or whether they may signal or even substitute each other. In particular, when only one of the two benchmarks is known, individuals can make inferences about the other benchmark and adjust their beliefs and actions accordingly. For instance, a well-known social norm can indicate what kind of behavior another person expects in a given situation. Take as an example a place where a tipping norm is on average 15%. In such an environment, one might generally believe a waiter to expect, on average, a 15% tip (as discussed in Charness and Dufwenberg, 2006).[3] Conversely, a disclosed expectation of another party will signal common conduct, especially in unfamiliar environments.[4] Hence, the experimental manipulation of just one of these two benchmarks will not be sufficient to separate a direct effect of a respective benchmark from an indirect effect arising as it signals the other one. Moreover, in a hypothetical case when a given benchmark does not matter for decision makers per se, one could still observe an effect of *information* about this benchmark on choices since it can signal the other (actually relevant) benchmark.

First, we formalize the claim that descriptive social norms and individual expectations can mutually signal each other. In the model, an agent who has to take an action, obtains a noisy signal about the population mean of this action (the descriptive social

[1] Also injunctive social norms (perception of what most others approve) can influence individual behavior (see Cialdini et al., 1990; Krupka and Weber, 2013). Since guilt aversion is mainly studied through subjects' expectations about the actual behavior, the expected confounding correlation between social norms and expectations, as discussed in this paper, is more relevant for descriptive social norms. Therefore, we deliberately focus on the conformity to the descriptive (and not injunctive) social norms.

[2] See Bardsley and Sausgruber (2005), Krupka and Weber (2009), Bicchieri and Xiao (2009), Köbis et al. (2015) for norm conformity, and Dufwenberg and Gneezy (2000), Charness and Dufwenberg (2006), Reuben et al. (2009), Di Bartolomeo et al. (2019), Mischkowski et al. (2019) for guilt aversion. Note, however, that Ellingsen et al. (2010) do not find evidence in line with guilt aversion (see Khalmetski et al. (2015) for a possible reconciliation).

[3] Such causal link between behavior and expectations was exploited by Balafoutas and Sutter (2017) to study guilt aversion in a dictator game. They informed dictators about the transfers received by their recipients in the past, thus inducing particular second-order beliefs of dictators about what their recipient might expect from them.

[4] See Sliwka (2007), Friebel and Schnedler (2011), Van der Weele (2009), and Bénabou and Tirole (2012) for a theoretical analysis and Danilov and Sliwka (2017) for corresponding experimental evidence that the revealed choices of informed principals might signal social norms to the agents.

norm). We show that:

1. If the agent is a norm complier but not guilt averse then under uncertainty about the social norm she reacts to disclosed expectations of others affected by her choice even though she does not care about these others' expectations *per se.*

2. Vice versa, if an agent is guilt averse but not a norm complier, then revealed information about the norm will affect her behavior even though she does not directly care about compliance to the social norm.

The reason for this behavior is simple: When all agents are uncertain about the actions taken by others in the population and receive private signals, any information on behavior in the population will also yield information about the belief of the affected party. At the same time, any information about the expectation of the affected party will also yield information about the population norm. In other words, norm compliance generates seemingly guilt averse behavior when there is uncertainty about the norm. Vice versa, guilt aversion produces seemingly norm-compliant behavior when there is uncertainty about the other player's expectation. It is then conceivable that previous results showing evidence for guilt aversion (i.e. reactions to disclosed expectations) may in fact be driven by norm compliance rather than guilt aversion. Similarly, previous results that have shown evidence for norm compliance (i.e., reactions to disclosed behavior by others in the same situation) may have also been driven by guilt aversion rather than norm compliance.

To study the cross-signaling effect between expectations and descriptive social norms empirically we conduct two experiments. Our experiments build on one-shot dictator games. In the first experiment, we study the behavior of dictators when *both* the expectation of the recipient and the average transfer of other dictators (from a previous experimental stage) are known before making transfers. Providing dictators with these two signals, which exogenously vary between subjects, allows us to separate the effects of norm compliance and guilt aversion. In particular, disclosed individual expectations should not convey to the dictator substantial additional information about the social norm as long as he simultaneously observes the actual average transfer, which is a much more precise signal about the descriptive norm. The same applies to the effect of information about the average transfer, which does not provide to the dictator any additional information about the recipient's expectation as long as the latter has been already disclosed. Hence, any possible effect of the recipient's expectation (or the average transfer) that we might observe is likely to be attributed to the guilt aversion (or norm compliance), since the cross-signaling effect is eliminated.

In the second experiment, we test the relationship between the guilt aversion and the compliance to a descriptive social norm from a different perspective. This time, the dictators are not informed about the behavior of other dictators but observe two benchmarks based only on recipients' expectations – the expectation of their own recipient and the expectation of an unrelated recipient. These two expectations transmit equally precise signals of a social norm, yet only the expectation of dictator's own recipient is relevant for guilt aversion. Therefore, the more similar is the effect of the two expectations, the more likely it is that they affect dictators' behavior because of the desire to comply with the social norm rather than guilt aversion. Whereas the larger the effect of the own recipient's expectation relative to the expectation of a random recipient, the more prevalent is guilt aversion.

To sum up, we exploit exogenous variation in the information disclosed to the dictator to disentangle the effects of norm conformity and guilt aversion from each other, and to assess the cross-signaling effects between descriptive norms and expectations of others. Notably, in our experiments we elicit recipients' expectations in an incentive-compatible manner, avoid deception, and control for dictators' social preferences.

We find that (i) descriptive social norms have a substantial positive effect on dictators' transfers and (ii) also the disclosed recipient's expectation affects dictators' behavior, but in a self-serving manner: A low expectation of the recipient is used as an excuse to lower the transfer, whereas a high expectation is rather ignored. The data from our second experiment, where dictators learn the expectation of an unrelated recipient instead of the descriptive norm, confirm the effect of the matched recipient's expectation observed in the first experiment. Yet, we find little evidence for a norm-signaling effect of the information on the random recipient's beliefs.

To the best of our knowledge, our study is the first aiming to disentangle guilt aversion from compliance to descriptive social norms by exploiting exogenous variation of information about *both* behavior and expectations of others.[5] There has been some work comparing compliance to injunctive social norms (i.e., commonly shared beliefs about socially appropriate behavior) and guilt aversion - but these papers do not control for potential cross-signaling effects between norms and individual expectations. Krupka et al. (2017) argue that ex-ante informal agreements affect behavior in symmetric games through changes in injunctive norms as well as by changing the expectations of the respective other party. The authors estimate structural models using information about injunctive norms elicited in a second experiment, and measuring the effect of guilt aversion through self-reported second-order beliefs. Hauge (2016) studies the effects of injunctive norms and recipients' expectations in the dictator game but does not display both benchmarks within the same treatment. In turn, results may still be driven by cross-signaling effects between norms and expectations.

Related to our findings on the non-monotonic effect of induced second-order beliefs,

[5]Exogenous variation of information about peer behavior was used to study the role of norm compliance in Bardsley and Sausgruber (2005), Bicchieri and Xiao (2009), Servátka (2009), Krupka and Weber (2009) and Bicchieri et al. (2021). Exogenous variation of information about others' expectations was used to study guilt aversion in Reuben et al. (2009), Ellingsen et al. (2010), Khalmetski et al. (2015), Bellemare et al. (2017), Bellemare et al. (2018) and Morell (2019).

some recent studies also show that the strength of guilt aversion depends on the "reasonableness" of expectations. Balafoutas and Fornwagner (2017) study the effect of exogenously varied recipient expectations in a dictator game. They find that the effect of expectations on giving even turns negative for a substantial share of dictators if these expectations exceed a certain level (which in turn varies across dictators). Similarly, Morell (2019) shows that very high expectations of recipients in a dictator game tend to be ignored by dictators. Pelligra et al. (2016) provide experimental evidence that the returned amounts of trustees in a trust game are not affected by reported expectations of the experimenter if these expectations exceed trustees' prior (i.e., unconditional) return amounts. A similar self-serving bias with respect to revealed *average* expectations has been reported in Charness et al. (2019). They find that the trustee does not react to the revealed average expectations of other trustees if these expectations imply a more cooperative behavior than her prior belief (yet, the effect is observed if the resulting belief adjustment is towards less cooperative behavior).

A number of other studies show that the effect of expectations of others is context-dependent. Vanberg (2008) finds that dictators are not sensitive to recipient expectations induced by promises of another dictator. Similarly, Khalmetski (2016) suggests that subjects tend to ignore higher expectations of another player if the latter are based on a different source of information than those of the decision maker. Di Bartolomeo et al. (2019) further find that the inclination to keep one's own promise is orthogonal to the second-order belief of the dictator, and is rather driven by the desire to keep the promise per se. Khalmetski et al. (2015) show that the aggregate effect of recipient's expectations may be statistically indistinguishable from zero if one does not control for the inclination to exceed others' expectations. d'Adda et al. (2019) show that a larger variance of exogenously disclosed beliefs of others about the injunctive norm results in a more dispersed distribution of transfers in the dictator game. Morell (2019) provides evidence that guilt aversion is stronger under shared group identity. The experiment of Charness and Rabin (2005) demonstrates that expressing a preference for being treated nicely by others after being selfish can backfire.[6]

The remainder of the paper is organized as follows. Section 2 provides a model formalizing mutual signaling of norms and expectations. Section 3 describes the experimental design, hypotheses and results of Experiment 1. Section 4 presents Experiment 2. Section 5 provides discussion and concludes.

[6]On a more general level, Dana et al. (2007), Ockenfels and Werner (2012), Exley (2016) and Bicchieri et al. (2018), among others, show that individuals can use also other motives, such as risk, uncertainty or lack of knowledge, as a self-serving excuse in charitable giving.

2 Model

2.1 Baseline setting

Consider an agent i who takes a personally costly action $a_i \in [0,1]$ which increases the utility of another agent j. The agent is potentially guilt averse and may have a preference for norm compliance such that her utility function is

$$u\left(a_i, E_j, N, \theta_{Gi}, \theta_{Ni}\right)$$

where $E_j = E_j\left[a_i\right]$ is the expectation of agent j about i's action, and N is a descriptive social norm. The social norm N is equal to the population mean of the chosen action $E\left[a\right]$, which is unknown to the agent. A more general assumption consistent with our results in this section is that the social norm is any statistic that is positively correlated with $N = E[a]$ (such as the mode). The agent's type is determined by $\theta_{Gi}, \theta_{Ni} \in \mathbb{R}_0^+$, where θ_{Gi} measures the degree of guilt aversion and θ_{Ni} is the propensity for norm compliance.

We furthermore assume that a higher action a_i leads to larger material costs, yet decreases psychological costs from falling behind j's expectations or the social norm (with the latter effects becoming dominant once i's sensitivity to guilt or norm compliance is sufficiently high):

- $\frac{\partial u}{\partial a_i} < 0$ for $\theta_{G_i}, \theta_{N_i} = 0$,
- $\frac{\partial u}{\partial a_i} > 0$ for $a_i < E_j$ if θ_{Gi} is larger than some cutoff $\bar{\theta}_G$,
- $\frac{\partial u}{\partial a_i} > 0$ for $a_i < N$ if θ_{Ni} is larger than some cutoff $\bar{\theta}_N$,
- $\frac{\partial u}{\partial a_i} < 0$ for $a_i > \max\left\{E_j, N\right\}$.

An example for such a utility function is

$$u\left(a_i, E_j, N, \theta_{Gi}, \theta_{Ni}\right) = K - a_i - \theta_{Gi} \max\left\{0, E_j - a_i\right\} - \theta_{Ni} \max\left\{0, N - a_i\right\}$$

where K is some parameter, which, for instance, nests a standard linear guilt aversion model (if $\theta_{Ni} = 0$) or a simple model of norm compliance (if $\theta_{G_i} = 0$). But the assumptions also allow for non-linear psychological costs of deviation from expectations or norm violation.

We now show that when there is uncertainty about either j's expectation E_j or the norm N the following holds, respectively:

(i) When an agent i learns a signal about j's expectation E_j but is *not* guilt averse (i.e., $\theta_{Gi} = 0$), then a_i is still increasing in this signal if i is sufficiently norm compliant.

(ii) When an agent i learns a signal about the norm N but is *not* norm compliant (i.e., $\theta_{Ni} = 0$), then a_i is still increasing in this signal if i is sufficiently guilt averse.

In other words, norm compliance generates guilt averse behavior when there is uncertainty about the norm. And vice versa, guilt aversion produces norm-compliant behavior when there is uncertainty about the other player's expectation. The argument is formalized in the subsequent sections.

2.2 Information structure

We consider the following information structure. Assume that agents ex-ante believe that N is distributed on $[0,1]$ according to a cumulative distribution function $G(N)$ with probability density $g(N)$.[7] Every agent i gets a noisy signal $s_i \in S \subset \mathbb{R}$ about the norm, which is independently and identically distributed according to a cumulative distribution function $F(s|N)$ with probability density $f(s|N)$. These signals can be interpreted, for instance, as some prior private knowledge about the strategic setting. All the distribution and density functions are assumed to be continuously differentiable in all arguments. We also assume that $F(s|N)$ satisfies the strict monotone likelihood ratio property (MLRP; Milgrom, 1981), i.e.

$$\forall s \in S, N'' > N' : \frac{\partial}{\partial s} \frac{f(s|N'')}{f(s|N')} > 0. \tag{1}$$

The MLRP implies that a higher norm leads to stochastically higher signals (in the sense of first-order stochastic dominance), while a higher signal leads to a higher conditional distribution of the norm. This is shown in the following auxiliary lemma.

Lemma 1 *For any $s \in S$ and $N \in [0,1]$ we have*

$$\frac{\partial F(s|N)}{\partial N} < 0, \tag{2}$$

$$\frac{\partial G(N|s)}{\partial s} < 0. \tag{3}$$

Proof: See Appendix A. ■

Denote by E_{Ni} the expectation of agent i about the norm after observing s_i, i.e.

$$E_{Ni} = E\left[N|s_i\right].$$

Lemma 1 straightforwardly implies that a higher private signal about the norm leads to a higher own expectation of the norm.

[7] The same notation is used for conditional distribution/density functions.

Corollary 1 *The expectation about the norm E_{Ni} strictly increases with s_i.*

Proof: Using integration by parts, we obtain:

$$E_{Ni} = E[N|s_i] = \int_0^1 Ng(N|s_i)dN = 1 - \int_0^1 G(N|s_i)dN.$$

Then, the claim follows by Lemma 1. ∎

Next, let us consider how the revealed information about the expectation of another agent affects one's own beliefs about the norm, and vice versa.

2.3 The effect of information about the expectation of another agent

Assume that agent i now additionally receives information on $j's$ beliefs about her own choice a_i. To be specific, let us assume that she directly observes $j's$ expectation $E_j = E_j\left[a_i|s_j\right]$ while the norm N remains unknown to both agents. As j does not have further information on i besides s_j this expectation must be equal to

$$E_j = E_j\left[a_i|s_j\right] = E_j\left[a|s_j\right] = E\left[N|s_j\right] = E_{Nj}. \tag{4}$$

Note that since E_{Nj} is a continuous and strictly increasing function of s_j by Corollary 1, agent i can perfectly infer s_j from observing E_{Nj} using the inverse function which we here denote as $s_j(E_{Nj})$. Then, agent i's posterior belief about the norm after observing both s_i and E_j is

$$E_i[N|s_i, E_j] = E_i[N|s_i, s_j] = \int_0^1 Ng(N|s_i, s_j)dN = 1 - \int_0^1 G(N|s_i, s_j)dN, \tag{5}$$

where to derive the last term we used integration by parts. Since s_i and s_j are independently distributed conditional on the norm N, the result of 1 still holds with respect to $G(N|s_i, s_j)$, i.e., $\frac{\partial G(N|s_i,s_j)}{\partial s_j} < 0$.[8] This together with (5) yields

$$\frac{\partial E_i[N|s_i, E_j]}{\partial s_j} > 0. \tag{6}$$

Finally, by the chain rule we obtain

$$\frac{\partial E_i[N|s_i, E_j]}{\partial E_j} = \frac{\partial E_i[N|s_i, E_{Nj}]}{\partial E_{Nj}} = \frac{\partial E_i[N|s_i, E_{Nj}]}{\partial s_j}\frac{\partial s_j(E_{Nj})}{\partial E_{Nj}} > 0, \tag{7}$$

where the first equality is by (4), while the inequality follows from (6) and Corollary 1. If the agent is not guilt averse ($\theta_{Gi} = 0$) but $\theta_{Ni} > \bar{\theta}_N$, she will thus choose

[8] In particular, the MLRP also holds for the conditional distribution $F(s_j|N, s_i)$ since

$$\frac{\partial}{\partial s_j}\frac{f(s_j|N'')}{f(s_j|N')} = \frac{\partial}{\partial s_j}\frac{f(s_j|N'', s_i)}{f(s_j|N', s_i)}.$$

Hence, the arguments in the proof of Lemma 1 continue to hold.

$a_i = E_i[N|s_i, E_j]$ which is strictly increasing in E_j by (7).

Hence, we have shown the following result:

Proposition 1 *Under uncertainty about the norm, observing a higher expectation of another agent leads to a higher posterior belief about the norm. In turn, even if the agent is not guilt averse ($\theta_{Gi} = 0$), higher revealed expectations of the other agent lead to a higher action, i.e.,*

$$\frac{\partial a_i}{\partial E_j} > 0$$

if the agent is sufficiently norm compliant ($\theta_{Ni} > \bar{\theta}_N$).

Proposition 1 thus implies that if an agent is a norm complier but not guilt averse then under uncertainty about the social norm she should react to disclosed expectations of others even though she does not care about these expectations *per se*.

2.4 The effect of information about the social norm

Let us show that the private signal s_i about the social norm also affects i's second-order beliefs about E_j in the case that the latter is unknown to i. The second-order belief of i about E_j is

$$\begin{aligned}
E_i[E_j] &= E_i[E_j|s_i] = E_i[E_{Nj}|s_i] = E_i\left[\int_0^1 Ng(N|s_j)dN \middle| s_i\right] \\
&= \int_S \left(\int_0^1 Ng(N|s_j)dN\right) f(s_j|s_i)ds_j \qquad (8) \\
&= \int_S \left(1 - \int_0^1 G(N|s_j)dN\right) f(s_j|s_i)ds_j \\
&= 1 - \int_S \left(\int_0^1 G(N|s_j)dN\right) f(s_j|s_i)ds_j \\
&= 1 - \int_0^1 G(N|\bar{s})dN + \int_S \int_0^1 \frac{\partial G(N|s_j)}{\partial s_j} dN F(s_j|s_i)ds_j, \qquad (9)
\end{aligned}$$

where we used integration by parts to obtain the third and the fifth equalities, and $\bar{s}$ is the upper bound of S. Consequently,

$$\frac{\partial E_i[E_j]}{\partial s_i} = \int_S \int_0^1 \frac{\partial G(N|s_j)}{\partial s_j} dN \frac{\partial F(s_j|s_i)}{\partial s_i} ds_j. \qquad (10)$$

At the same time, using the law of total probability and integration by parts, we obtain

$$\begin{aligned}
F(s_j|s_i) &= \int_0^1 F(s_j|N, s_i)g(N|s_i)dN \\
&= \int_0^1 F(s_j|N)g(N|s_i)dN \\
&= F(s_j|1) - \int_0^1 \frac{\partial F(s_j|N)}{\partial N} G(N|s_i)dN.
\end{aligned}$$

This together with Lemma 1 implies

$$\frac{\partial F(s_j|s_i)}{\partial s_i} < 0.$$

Substituting this into (10) and taking into account $\frac{\partial G(N|s_j)}{\partial s_j} < 0$ by Lemma 1, we finally have

$$\frac{\partial E_i[E_j]}{\partial s_i} > 0. \tag{11}$$

If the agent is not norm compliant ($\theta_{Ni} = 0$) but $\theta_{Gi} > \bar{\theta}_G$, she will thus choose $a_i = E_i[E_j]$ which is strictly increasing in s_i by (11). Hence, we obtain the following result.

Proposition 2 *Under uncertainty about the expectation of another agent, a higher signal about the norm leads to a higher second-order belief about this expectation. In turn, even if the agent is not norm compliant ($\theta_{Ni} = 0$), a higher signal about the norm leads to a higher action, i.e.,*

$$\frac{\partial a_i}{\partial s_i} > 0$$

if the agent is sufficiently guilt averse ($\theta_{Gi} > \bar{\theta}_G$).

Analogously to Proposition 1, Proposition 2 implies that if an agent is purely guilt averse, then revealed information about the norm may change her behavior even though she might not directly care about compliance to social norms.

3 Experiment 1: Norm conformity and guilt aversion

3.1 Design and procedures

Experiment 1 consists of two sequentially conducted identical dictator games (henceforth *Game 1* and *Game 2*). In each dictator game the dictator had to allocate €14 between herself and a randomly assigned recipient.

In the beginning, all participants received the instructions for Game 1 and learned about their roles. The roles (either of dictator or recipient) were assigned randomly and remained unchanged until the end of the experiment. After reading the instructions for Game 1, recipients had to estimate the average dictator transfer in their experimental session and dictators were asked to state their beliefs about the recipient's guess of the average transfer, i.e., to submit their second-order beliefs. Subjects received €5 for their answers that deviated from the true value by no more than €0.15. After Game 1 was finished, subjects were informed about whether their guesses earned the bonus.[9]

[9]This timing allows to better control for subjects' beliefs regarding the total payoff from the experiment in the regression analysis (by including the bonus dummy as independent variable). We asked recipients to guess the average behavior of dictators, rather than the individual transfer of their matched dictator, to avoid potential bias in reported beliefs in that recipients hedge their experimental

After Game 1, the instructions for Game 2 were distributed and each dictator was matched with a new recipient. Each recipient was then asked whether she agrees that her expectation, elicited at the beginning of Game 1, can be transmitted to the respective dictator in Game 2. If the recipient agreed, she obtained an additional payoff of €2.50 in Game 2. This procedure was used to reduce experimental deception by omission: Recipients became fully aware and in control of their belief disclosures to dictators, while – at the same time – they couldn't strategically distort their guesses at the beginning of the experiment (see Khalmetski et al., 2015, for a discussion). Out of 124 recipients, 121 agreed to transmit their guesses to dictators so that we obtained 121 observations for the analysis, while avoiding substantial selection effects. Also in Game 2, dictators had to decide how to allocate €14 between themselves and their newly matched recipient. Before this decision, they could see two pieces of information: (i) the average transfer that *the other dictators in the same session* made in Game 1 and (ii) the guess of the matched recipient about the average transfer of the dictators (if the permission to transmit the guess was granted by the respective recipient). The information was presented to the dictators in a random order.

We focus on the average action of others as a signal of the descriptive social norm (in both the theory and experiment) for two reasons. First, the average action is a simple statistic which is arguably correlated with the descriptive social norm, i.e., common behavior (Bohnet and Zeckhauser, 2004). It is in particular useful under a continuous action space when no single actions might be chosen by a clear majority of the population. Second, the average action is similar to the notion of the individual expectation as used in the guilt aversion literature allowing to measure both expectations and norms in comparable manner (the realized average action vs. the expected average action).[10]

After dictators made their decision in Game 2, a random draw of three possible earnings (bonus for belief-elicitation questions, income from Game 1, income from Game 2) determined the final cash payment.[11]

The experiment was conducted in the Cologne Laboratory for Economic Research with 248 participants (mostly, students of the University of Cologne) in December 2015. Participants were recruited with ORSEE (Greiner, 2015), and the experiments

income using their stated estimate (see Schlag et al., 2015, for a discussion). Note also that all subjects were asked to report just one type of beliefs (first-order beliefs for recipients and second-order beliefs for dictators.)

[10]The disclosure of the average behavior of others was used to study social conformity also in many other empirical studies, see, e.g., Fischbacher et al. (2001), Bohnet and Zeckhauser (2004), Schultz et al. (2007), and Allcott (2011). From the empirical perspective, if anything, this approach should weaken our results as compared to a situation when individuals are (also) informed about other moments of the distribution of others' behavior (econometrically, the norm is then measured with noise leading to downward attenuation bias).

[11]In order to make the earnings more equal between Game 1, Game 2 and belief-elicitation questions, we added a fixed additional payment of €7.50 in the belief-elicitation stage. Additionally, subjects received a fixed payment of €2.50 in both Game 1 and Game 2 (with an exception of recipients in Game 2, who were instead compensated with €2.50 for the belief transmission, which then equalized their initial endowment with that of the dictators).

were computerized with z-Tree (Fischbacher, 2007). We ran a total of 16 sessions.[12] The instructions were distributed on paper and can be found in Appendix C. The average earning was €8.9, while the experiment lasted for about 45 minutes.

3.2 Research question

In Section 2 we formally show a mutual signaling effect: Revealed expectations of others and descriptive social norms can signal each other. Hence, in order to distinguish between guilt aversion and norm conformity, one needs to control for such cross-signaling effects. The most straightforward way to do this is to ensure that subjects learn information about both, the descriptive norm and the recipient's expectation. In this case, the cross-signaling effects, which rely on incomplete information, are excluded. In particular, a revealed expectation of another subject, being itself a noisy signal about the norm, shouldn't significantly influence the belief about the norm if subjects already obtained a sufficiently precise signal about the norm, and vice versa. In turn, the natural sampling variation of the average transfers of others and the recipient's expectation across the sessions and dictators helps us identify the causal effects of descriptive norms and guilt aversion on behavior.

In sum, our experimental design allows us to disentangle the direct effect of the recipient's expectation from that of the observed behavior of others on dictator transfers, and to test whether these effects are statistically significant. Our first experimental hypothesis is that, controlling for the disclosed recipient's expectation, dictator transfers increase with the observed average behavior of other dictators (norm conformity). Our second experimental hypothesis is that, controlling for the observed behavior of others, dictator transfers increase with the disclosed recipient's expectation (guilt aversion).

3.3 Results

As a first step it is important to see whether the heterogeneity in individual transfers and recipients' expectations generated sufficient exogenous variation in the information displayed to dictators.

The average transfer of others displayed to dictators in Game 2 indeed varied between €0.86 and €4.82 (Mean = 3.12, SD = 0.92).[13] The transmitted recipient's expectation varied between €0 and €13.85 (Mean = 4.17, SD = 2.98). Around 90% of the expectations were below or equal to the half of the pie.[14] Figure 1 gives an overview of the distribution of both benchmarks.

[12]Each session included either 14 or 16 subjects. The variation was caused by different attendance rates.

[13]Note that this is well in line with the natural variation arising from an ex-post Monte Carlo simulation on the basis of our data, see Appendix B.

[14]The average dictator's second-order belief was €4.32 (SD = 3.48), with 89.5% of the beliefs being below or equal to the half of the pie.

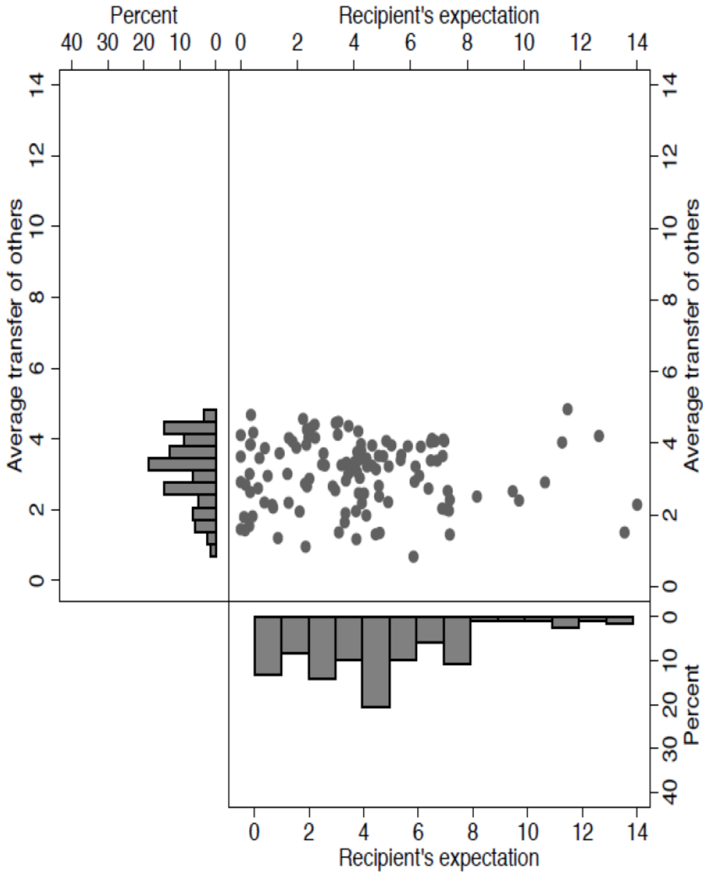


Figure 1: Distribution of descriptive norms and recipients' expectations in Experiment 1.

Table 1: Effect of descriptive norm and recipient's expectation on transfers in Experiment 1.

	(1) All	(2) Exp.∈ [€0.86, €4.82]
Unconditional transfer (Game 1)	0.655***	0.563***
	(0.042)	(0.063)
Norm (average transfer)	0.383***	0.558***
	(0.141)	(0.189)
Recipient's expectation	0.013	0.362**
	(0.045)	(0.169)
Observations	121	64
Pseudo R^2	0.241	0.214

Tobit regressions; marginal effects reported; the bonus for correct beliefs and the information order are controlled for; robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

The average transfer in Game 2 was €2.56 (SD = 2.32). This is significantly lower than in Game 1 ($p < 0.01$, Wilcoxon signed-rank test). Approximately half of the dictators in our sample did not change their allocation decisions in Game 2 relative to Game 1 (50.8%). Among those dictators who decreased transfers (36.3%), a median change in transfers was 33.33%. Around 13% of subjects transferred more in Game 2 as compared to Game 1 (the median increase was 41.67%).

Exploiting the exogenous variation in descriptive norms and recipient expectations, we can now estimate their causal effect on dictators' transfers. Table 1 shows the regression results (marginal effects from Tobit estimations) with the dictator transfer in Game 2 as the dependent variable and the descriptive norm and the expectation of the matched recipient as independent variables. We also control for the dictator's own transfer in Game 1 as a measure for general social preferences, whether the dictator has earned a bonus for the initial belief-elicitation questions (in which case her expected payoff from the experiment is higher, all else equal), and for the order of displayed information (i.e., whether the recipient's expectation was displayed before or after the norm).[15]

We find strong evidence in favor of norm conformity: As column (1) shows, the information about the descriptive social norm has a sizeable and significant effect on dictators' transfers. An increase in the displayed norm by €1 increases dictators' transfers by €0.38.

At the same time, in model (1) the recipient's expectation does not have a significant effect on the dictator's transfer. This might suggest that guilt aversion plays no role when individuals have information about the social norm, while recipients' expectations may hypothetically affect dictators' behavior only as they signal information about the norm. Yet, a closer inspection of the data delivers a more nuanced picture.

First of all, it is important to note that some previous experiments (Ellingsen et al., 2010; Khalmetski et al., 2015) in which dictators did not have information about the norm also did not detect aggregate effect of recipients' (exogenously disclosed) expectations. One potential reason is that guilt aversion matters only if the expectations of the recipient are deemed "acceptable" or "legitimate". Balafoutas and Fornwagner (2017) show that expectations raise transfers in the dictator game only if they do not exceed a certain (individual specific) level. Moreover, "unreasonably" high expectations beyond that level may even be "punished" with lower transfers. In our data, the recipient's expectations vary across a much larger interval (i.e., between €0 and €13.85) than the norm (between €0.86 and €4.82). Hence, it is possible that observed expectations are more likely to appear in an "unreasonable" range than the norm. To account for this, we reduce the sample to the observations where the expectations lie in the interval spanned by the norm (i.e., between €0.86 and €4.82).[16] As shown in column 2 of Table

[15]The coefficients on the latter two variables are not significant in this and subsequent regressions and are omitted in the tables.

[16]Note that the expectation of the recipient is a purely exogenous variable from the perspective of

1, the coefficient on the recipient's expectation then becomes sizeable and significant.

This leads to two questions: First, can we say more about which expectations dictators deem acceptable? And second, how do dictators react to expectations beyond the acceptable range? In particular, there may be different benchmarks in our context to which an expectation could be compared by dictators. These potential benchmarks include the equal split, the dictator's unconditional transfer in Game 1, the dictator's second-order belief about the expectation of the recipient, and the displayed descriptive social norm. Table 2 shows regressions of the form

$$\begin{aligned} transfer_i &= \alpha + \beta \cdot transfer_i^0 + \gamma \cdot norm_i + \delta \cdot expectation_i \\ &\quad + \phi \cdot (expectation_i - benchmark_i) \cdot I_{\{expectation_i \geq benchmark_i\}} \end{aligned}$$

where $transfer_i$ denotes transfer in Game 2, $transfer_i^0$ is the (unconditional) transfer under no information in Game 1, and $I_{\{expectation_i \geq benchmark_i\}}$ is a dummy variable indicating whether the expectation exceeds the respective benchmark. In this way, we estimate piece-wise linear and continuous reaction functions (with respect to a change in the recipient's expectation). Here, ϕ estimates a potential change in the slope of the reaction function at the respective benchmark.[17] We compare different potential benchmarks, i.e., the previous transfer in Game 1, the displayed descriptive norm, the equal split of 7, and the dictator's second-order belief about the recipient's expectation (SOB). Each model thus estimates whether the reaction function displays a kink at the relevant benchmark, i.e., whether its slope is significantly different above as compared to below the benchmark.

Table 2 shows the results of the regression analysis. As model (1) indicates, there is a kink in the dictator's reaction function at her own prior transfer choice in Game 1 (i.e., the transfer chosen before receiving information about the norm and the recipient's expectation). Below this threshold, transfers are increasing in the recipient's expectations, but above this threshold they are strictly decreasing. As the Wald test confirms, the estimated slope beyond the kink of $0.285 - 0.454 = -0.169$ is significantly smaller than zero ($p < 0.001$). We find no evidence that there is a structural break at any of the other potential benchmarks.[18]

the dictator, and hence such sample restriction does not reduce the statistical validity of the regression analysis.

[17] Indeed, assume that the true data generating process for some dependent variable y and regressor x is of the form:

$$y_i = \begin{cases} \alpha + \beta_1 x_i + \gamma b_i + \varepsilon_i \text{ if } x_i < b_i, \\ \alpha + \beta_1 b_i + \beta_2 (x_i - b_i) + \gamma b_i + \varepsilon_i \text{ if } x_i \geq b_i, \end{cases}$$

with $\beta_2 \neq \beta_1$ and b_i is some individual benchmark. Thus, y as a function of x has a kink at the benchmark (while γ captures the independent effect of b_i). One can easily verify that this expression for y_i is equivalent to

$$y_i = \alpha + \beta_1 x_i + (\beta_2 - \beta_1)(x_i - b_i) I_{\{x_i \geq b_i\}} + \gamma b_i + \varepsilon_i$$

so that the resulting coefficient on $(x_i - b_i) I_{\{x_i \geq b_i\}}$ measures the change in the slope at the benchmark.

[18] Estimating models (1) and (2) on the restricted sample with expectations from €0.86 to €4.82

Table 2: Reference standards for the observed expectation in Experiment 1.

	(1)	(2)	(3)	(4)
Unconditional transfer (Game 1)	0.398***	0.656***	0.656***	0.642***
	(0.0909)	(0.0410)	(0.0410)	(0.0470)
Norm (average transfer)	0.386***	0.360*	0.388***	0.376***
	(0.134)	(0.195)	(0.141)	(0.141)
Recipient's expectation	0.285***	0.0418	0.0669	0.0593
	(0.0917)	(0.150)	(0.0666)	(0.0755)
(Recip. expectation − Uncond. transfer)	-0.454***			
$\times\ I_{\{Recip.\ expectation\ \geq\ Uncond. transfer\}}$	(0.107)			
(Recip. expectation − Norm)		-0.039		
$\times\ I_{\{Recip.\ expectation\ \geq\ Norm\}}$		(0.175)		
(Recip. expectation − 7)			-0.196*	
$\times\ I_{\{Recip.\ expectation\ \geq\ 7\}}$			(0.115)	
(Recip. expectation - SOB)				-0.0679
$\times\ I_{\{Recip.\ expectation\ \geq\ SOB\}}$				(0.0722)
Observations	121	121	121	121
Pseudo R^2	0.272	0.241	0.244	0.242

Tobit regressions; marginal effects reported; the bonus for correct beliefs and the information order are controlled for; robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

To illustrate the structural break at the unconditional expectation estimated in model (1) of Table 2 graphically, we regress the difference in transfers between Game 2 and Game 1 (i.e., the update in transfers between the games) on the dummy variables measuring how much the observed information (either about the norm or the recipient's expectation) deviates from the dictator's unconditional transfer in Game 1. Thus, we run the following specifications allowing for a non-linear reaction to the novel information contained in the norm and the recipient's expectation:

$$\begin{aligned}\Delta transfer_i \quad = \quad & \alpha + \gamma_1 I_{\{\Delta_i^{norm}<-3\}} + \gamma_2 I_{\{\Delta_i^{norm}\in[-3,-1]\}} + \gamma_3 I_{\{\Delta_i^{norm}\in[1,3]\}} \\ & + \gamma_4 I_{\{\Delta_i^{norm}>3\}} + \delta_1 I_{\{\Delta_i^{exp.}<-3\}} + \delta_2 I_{\{\Delta_i^{exp.}\in[-3,-1]\}} \\ & + \delta_3 I_{\{\Delta_i^{exp.}\in[1,3]\}} + \delta_4 I_{\{\Delta_i^{exp.}>3\}} + \varepsilon_i\end{aligned}$$

where $\Delta_i^{norm} = norm_i - transfer_i^0$ and $\Delta_i^{exp.} = expectation_i - transfer_i^0$. The interval $(-1, 1)$ is taken as a baseline. Hence, the coefficients γ_k (δ_k), $k = 1, ..., 4$, reflect how much the dictator updates her transfer in Game 2 after learning that her prior transfer deviated from the norm (expectation) by the corresponding amount, relative to the case where the observed norm (expectation) was close to the prior transfer. Figure 2 plots the coefficients for the considered intervals and their 95% confidence bands.

The right panel shows the dictator's reaction to information about the recipient's

delivers very similar results.

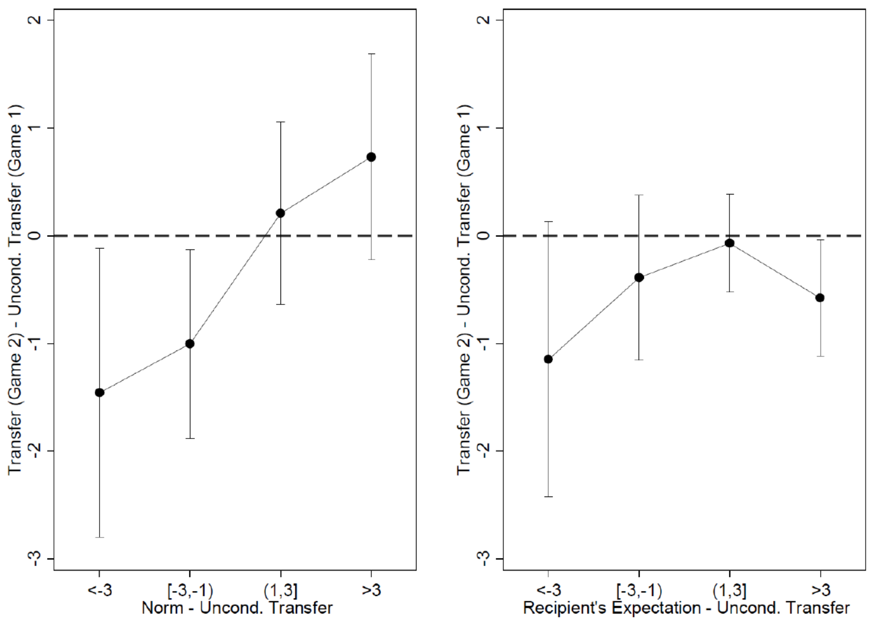


Figure 2: Reaction to information about the norm (left panel) and the recipient's expectation (right panel) in Experiment 1.

expectation that differs from her own previous transfer. It confirms the pattern detected in our piece-wise linear regression: Dictators reduce their transfers when learning that their recipient expects less than the previously given amount, but do not increase their transfers if the recipient expects more (even "punishing" too high expectations with lower transfers). In contrast, dictators react monotonically to information about the norm as the left panel shows: They reduce their transfers when the norm is lower than their own prior transfer, and tend to increase their transfers if it is above.

Hence, a key difference in the reaction to the two benchmarks is that while very high expectations by the recipient tend to induce lower actions, a higher descriptive norm raises actions monotonically. A potential explanation for this difference is that it may be harder to ignore (or dismiss as 'unreasonable') the information on the descriptive norm as it reflects the behavior of (*multiple*) other dictators as compared to the recipient's expectation that mirrors a subjective perception of only *one* individual. Moreover, actual behavior might be perceived as a more objective, and hence less malleable benchmark relative to expectations (which can be 'right' or 'wrong').

In sum, when both the descriptive social norm and the expectation of the recipient are known, (i) dictator transfers are increasing in the social norm, (ii) dictators positively react to the recipient's expectation when it falls below the dictator's unconditional transfer, and (iii) dictators negatively react towards the expectations that exceed their unconditional transfers. In other words, the effects of the descriptive social norm and revealed recipient expectations on dictator transfers are the following: Information about the social norm pushes transfers both ways (above and below the prior transfer).

At the same time, information about the recipient's expectation is used in a self-serving manner to reduce one's own giving.

4 Experiment 2: Guilt aversion under uncertainty about the social norm

4.1 Design and procedures

The structure and procedures of Experiment 2 were identical to those of Experiment1 except for the information shown in Game 2.[19] In particular, subjects were again assigned to the roles of dictators and recipients. They participated in two sequential dictator games where dictators could distribute €14 between themselves and an anonymous recipient. In the same manner as in Experiment 1, recipients and dictators were asked about their first- and second-order beliefs, respectively, before the start of Game 1. Then, dictators made their decisions without receiving any other information in Game 1. In Game 2, dictators could again observe two pieces of information conditional on the consent of respective recipients. Like in Experiment 1, one piece of information was the expectation of the recipient who was matched to the dictator in Game 2. Unlike to Experiment 1, the second piece of information was not the average transfer of other dictators in Game 1, but the expectation of *another randomly selected recipient* (whose payoff did not depend on the dictator's decision in Game 2, and who also had not been matched to the dictator in Game 1). These two pieces of information were presented in a random order.

The experiment was conducted in the Cologne Laboratory for Economic Research with 256 participants (mostly, students of the University of Cologne) in September 2018. Participants were recruited with ORSEE (Greiner, 2015), and the experiments were computerized with z-Tree (Fischbacher, 2007). We ran a total of 16 sessions with 16 subjects each. Out of 128 recipients, 126 agreed to transmit their guesses to dictators. The average earning was €10.7, while the experiment lasted for about 45 minutes. The remaining procedures were identical to those of Experiment 1. The experimental instructions can be found in Appendix C.2.

4.2 Research question

In Experiment 2, dictators learn simultaneously the expectations of their matched recipient and of one unrelated recipient. Hence, in Experiment 2 we don't show a direct indicator of the descriptive social norm, but rather control for the norm-signaling

[19] The Cologne Laboratory for Economic Research increased the size of the show-up fee from €2.5 to €4 in the interim time between our Experiments 1 and 2. Therefore, we adjusted the payment for the belief transmission of the recipients in Game 2 and the fixed payments in Games 1 and 2 to €4. To equalize the expected payment between Game 1, Game 2 and the belief-elicitation stage, the total fixed payment for the latter was €9. The different levels of rewards for belief transmission did not affect the recipients' consent rates between Experiments 1 and 2.

effect, which is potentially embedded in disclosed recipient's expectations, indirectly. In particular, both expectations disclosed to the dictator transmit equally precise signals about the descriptive social norm, yet only the matched expectation is relevant for guilt aversion. Hence, the difference between the effects of these expectations should be indicative of the pure effect of guilt aversion net of norm-signaling. At the same time, the magnitude of the effect of the expectation of an unrelated recipient should reveal the norm-signaling effect generally embedded in disclosed expectations.

Accordingly, our first experimental hypothesis is that dictator transfers increase with the disclosed expectation of a random recipient (norm signaling). Our second experimental hypothesis is that the disclosed expectation of the matched recipient has a larger effect on dictator transfers than the disclosed expectation of a random recipient (guilt aversion).

4.3 Results

In general, dictators' transfers are very similar in Experiment 2 to those in Experiment 1. In Game 2, 62.5% of dictators did not change their allocation decision (relative to Game 1), 27.3% decreased their transfer and 10.2% increased it. The average transmitted recipient's expectation was €4.35 varying between €0 and €13.86 (SD = 3.34; 89.7% of the beliefs were below or equal to the half of the pie). The average dictator transfer was €3.07 (SD = 2.68) in Game 1 and €2.67 (SD = 2.64) in Game 2. The average dictator's second-order belief was €4.41 (SD = 2.84).

Table 3 replicates the basic specification from Table 1 for the data from Experiment 2. The main difference is that instead of the average transfer of others as a proxy for the social norm (which is now unobservable to dictators) we now include the random expectation (of an unrelated recipient) as the second key independent variable. The overall picture is consistent with Experiment 1: The effect of the matched recipient's expectation is not significant in the whole sample, yet significant and sizeable under the restriction on the range of expectations to the interval from €0.86 to €4.82, which is the range we used for the analysis of Experiment 1 (see column 2). Note that both random and matched expectations are restricted to ensure the comparability of the respective coefficients.

The effect of the random other recipient's expectation is not significantly different from zero (in either the full or the restricted sample). More importantly, it is significantly smaller than the effect of the matched expectation in the restricted sample (Wald test, $p = 0.049$). This suggests that the observed effect of the matched expectation cannot be reduced to a pure-norm signaling effect but there should be a direct effect of guilt aversion.

In a next step, we study a potential kink of the effect of guilt aversion at various benchmarks, similar to our analysis for Experiment 1. In addition to the previously considered benchmarks (except for the average behavior as it is not observable for

Table 3: Effect of observed recipients' expectations on transfers in Experiment 2.

	(1) All	(2) Exp. $\in$ [€0.86, €4.82]
Unconditional transfer	0.721^{***}	0.610^{***}
	(0.0410)	(0.0582)
Matched recipient's	-0.0243	0.550^{**}
	(0.0557)	(0.247)
Random recipient's	-0.00573	0.305
	(0.0375)	(0.188)
Observations	124	31
Pseudo R^2	0.213	0.428

Tobit regressions; marginal effects reported; the bonus for correct beliefs and the information order are controlled for; robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The sample size in model (2) was restricted to the data points where both observed expectations were from the interval [€0.86; €4.82].

Table 4: Effect of observed recipients' expectations on transfers in Experiment 2.

	(1)	(2)	(3)	(4)
Unconditional transfer	0.597^{***}	0.705^{***}	0.669^{***}	0.718^{***}
(Game 1)	(0.123)	(0.0439)	(0.0673)	(0.0416)
Matched recipient's	0.130^{**}	-0.0341	-0.0267	0.0132
expectation	(0.0659)	(0.0890)	(0.0949)	(0.0817)
Random recipient's	-0.0101	0.0641	0.0962	-0.0276
expectation	(0.0885)	(0.0547)	(0.111)	(0.0509)
(Matched expectation − Uncond. transfer)	-0.238^{*}			
$\times I_{\{Matched\ expectation\ \geq\ Uncond.\ transfer\}}$	(0.133)			
(Random expectation − Uncond. transfer)	-0.00768			
$\times I_{\{Random\ expectation\ \geq\ Uncond.\ transfer\}}$	(0.107)			
(Matched expectation - 7)		0.0213		
$\times I_{\{Matched\ expectation\ \geq\ 7\}}$		(0.168)		
(Random expectation - 7)		-0.193		
$\times I_{\{Random\ expectation\ \geq\ 7\}}$		(0.138)		
(Matched expectation - SOB)			-0.0171	
$\times I_{\{Matched\ expectation\ \geq\ SOB\}}$			(0.0921)	
(Random expectation - SOB)			-0.150	
$\times I_{\{Random\ expectation\ \geq\ SOB\}}$			(0.158)	
(Matched expectation - Random expectation)				-0.0592
$\times I_{\{Matched\ expectation\ \geq\ Random\ expectation\}}$				(0.0888)
Observations	124	124	124	124
Pseudo R^2	0.221	0.216	0.218	0.214

Tobit regressions; marginal effects reported; the bonus for correct beliefs and the information order are controlled for; robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

dictators in Experiment 2), we additionally estimate whether the effect of the matched expectation depends on whether it lies above or below the random expectation. In this way, we can study whether dictators are prone to a self-serving bias by comparing both expectations and choosing to conform to the lowest one.

The results are shown in Table 4. Again consistent with Experiment 1, dictators tend to conform to the expectation of the matched recipient if it is lower than the dictator's unconditional transfer, in which case the effect of the matched expectation is significantly positive (see column 1). At the same time, matched expectations exceeding the unconditional transfer do not increase transfers: The estimated slope beyond the kink, $0.130 - 0.238 = -0.108$, is not significantly different from 0 ($p = 0.255$, Wald test). The effect of the random expectation is still insignificant and does not interact with the level of unconditional transfer. Also in line with Experiment 1, there is no statistically significant kink at the equal split or dictator's own second-order belief for either matched or random expectations (see columns 2 and 3). Finally, we do not find evidence that dictators react less strong to the matched expectation if it exceeds the random expectation (see column 4). A potential reason is that it might be difficult for a dictator to find a compelling reason to dismiss the expectation of one's own recipient in favour of the expectation of an unrelated recipient.[20]

Figure 3 shows reactions to deviations of the novel information (about the two types of recipient expectations) from the unconditional transfer (analogously to Figure 2 for Experiment 1). The dependent variable is again the difference between the dictator's transfers in Game 2 and Game 1. The figure plots the coefficients and their 95% confidence bands. The reaction to the matched recipient's expectation (right panel) follows a similar pattern as in Experiment 1: Dictators use the information about the matched recipient's expectation in a self-serving manner, only to reduce their transfers. At the same time, dictators simply stick to their unconditional transfer if the recipient's expectation exceeds it. Compared to Experiment 1, there is less "punishment" of expectations exceeding the unconditional transfer under uncertainty about the norm. The graph on the left panel may be slightly suggestive that the information about a randomly chosen other recipient's expectation may serve as a signal of the social norm, as the update in transfers follows an increasing pattern. Yet, the effect is statistically not very pronounced (the coefficient on the "> 3" dummy is only significant at the 10% level).

[20]To further study a potential self-serving bias of conforming to the lowest expectation, we estimated whether the effect of the random (matched) expectation is stronger in a subsample where the random expectation is smaller (larger) than the matched expectation. However, we found that the effect of neither random nor matched expectation is amplified in these subsamples.

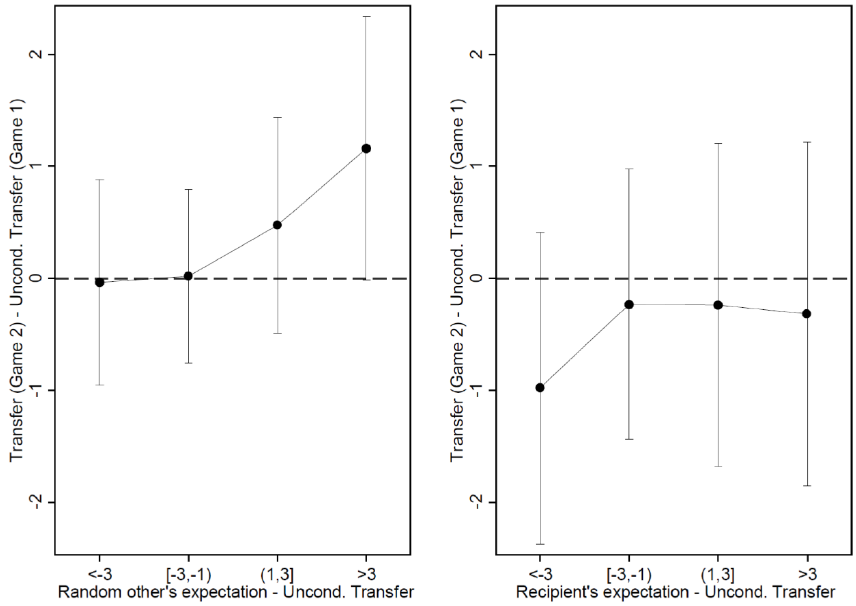


Figure 3: Reaction to information about the matched recipient's expectation (right panel) and the random other recipient's expectation (left panel) in Experiment 2.

5 Conclusion

Overall, our results point out that both norm conformity and guilt aversion are important in shaping individual decisions. In particular, we find that both (descriptive) social norms and revealed expectations of others affect behavior even if one controls for the fact that both of these types of information may mutually signal each other (Experiment 1). At the same time, the effect of the expectation has a kink at the unconditional transfer, which can be interpreted as dictators finding an 'excuse' not to comply to the recipient's expectation as far as the latter exceeds an 'appropriate' transfer level. Our results from Experiment 2 generally confirm that guilt aversion matters also under uncertainty about the descriptive norm, and this effect cannot be reduced to a pure norm-signaling effect. This is important from a methodological perspective as this validates the results from experiments using individual expectations without debriefing subjects about the descriptive norm.

Our results on non-monotonicity of guilt aversion comply with the previous related evidence of Pelligra et al. (2016), Khalmetski (2016) and Balafoutas and Fornwagner (2017), and motivate a refinement of the notion of guilt aversion formalized by Battigalli and Dufwenberg (2007). In particular, it might be reasonable to assume that not all expectations of other players equally matter for the utility of a (guilt averse) decision maker while those which exceed one's own counterfactual behavior (under prior beliefs) are downweighted. Our results could also refine the findings of Khalmetski et al. (2015) who showed that in order to measure the actual effect of guilt aversion one needs to control for the heterogeneity in belief-dependent preferences (such as eagerness to positively surprise others). The current study further complements it by showing that the heterogeneity in the revealed expectations should also be taken into account.

References

Allcott, H. (2011). Social norms and energy conservation. *Journal of public Economics 95*(9-10), 1082–1095.

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

Balafoutas, L. and M. Sutter (2017). On the nature of guilt aversion: Insights from a new methodology in the dictator game. *Journal of Behavioral and Experimental Finance 13*, 9–15.

Bardsley, N. and R. Sausgruber (2005). Conformity and reciprocity in public good provision. *Journal of Economic Psychology 26*(5), 664 – 681.

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

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

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

Bénabou, R. and J. Tirole (2012). Laws and norms. *IZA Discussion Paper 6290*.

Bicchieri, C., E. Dimant, and S. Sonderegger (2018). It's not a lie if you believe it: Lying and belief distortion under norm-uncertainty. *Philosophy, Politics and Economics (PPE) Working Papers, University of Pennsylvania*.

Bicchieri, C., E. Dimant, and E. Xiao (2021). Deviant or wrong? the effects of norm information on the efficacy of punishment. *Journal of Economic Behavior and Organization 188*, 209–235.

Bicchieri, C. and E. Xiao (2009). Do the right thing: but only if others do so. *Journal of Behavioral Decision Making 22*(2), 191–208.

Bohnet, I. and R. Zeckhauser (2004). Social comparisons in ultimatum bargaining. *Scandinavian Journal of Economics 106*(3), 495–510.

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

Charness, G., M. Naef, and A. Sontuoso (2019). Opportunistic conformism. *Journal of Economic Theory 180*, 100 – 134.

Charness, G. and M. Rabin (2005). Expressed preferences and behavior in experimental games. *Games and Economic Behavior 53*(2), 151–169.

Cialdini, R. B., R. R. Reno, and C. A. Kallgren (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology 58*(6), 1015.

d'Adda, G., M. Dufwenberg, F. Passarelli, and G. Tabellini (2019). Partial norms. Technical report.

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

Danilov, A. and D. Sliwka (2017). Can contracts signal social norms? experimental evidence. *Management Science 63*(2), 459–476.

Di Bartolomeo, G., M. Dufwenberg, S. Papa, and F. Passarelli (2019). Promises, expectations & causation. *Games and Economic Behavior 113*, 137 – 146.

Dufwenberg, M. and U. Gneezy (2000). Measuring beliefs in an experimental lost wallet game. *Games and economic Behavior 30*(2), 163–182.

Ellingsen, T., M. Johannesson, S. Tjøtta, and G. Torsvik (2010). Testing guilt aversion. *Games and Economic Behavior 68*(1), 95–107.

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

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

Fischbacher, U., S. Gächter, and E. Fehr (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economics letters 71*(3), 397–404.

Friebel, G. and W. Schnedler (2011). Team governance: Empowerment or hierarchical control. *Journal of Economic Behavior & Organization 78*(1), 1–13.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association 1*(1), 114–125.

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

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

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

Köbis, N. C., J.-W. Van Prooijen, F. Righetti, and P. A. Van Lange (2015). "who doesn't?" – the impact of descriptive norms on corruption. *PloS one 10*(6), e0131830.

Krupka, E. L., S. Leider, and M. Jiang (2017). A meeting of the minds: Informal agreements and social norms. *Management Science 63*(6), 1708–1729.

Krupka, E. L. and R. A. Weber (2009). The focusing and informational effects of norms on pro-social behavior. *Journal of Economic psychology 30*(3), 307–320.

Krupka, E. L. and R. A. Weber (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association 11*(3), 495–524.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics 12*(2), 380–391.

Mischkowski, D., R. Stone, and A. Stremitzer (2019). Promises, expectations, and social cooperation. *The Journal of Law and Economics 62*(4), 687–712.

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

Ockenfels, A. and P. Werner (2012). 'hiding behind a small cake'in a newspaper dictator game. *Journal of Economic Behavior & Organization 82*(1), 82–85.

Pelligra, V., T. Reggiani, and D. J. Zizzo (2016). Responding to (un) reasonable requests. Technical report.

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

Schlag, K. H., J. Tremewan, and J. J. Van der Weele (2015). A penny for your thoughts: A survey of methods for eliciting beliefs. *Experimental Economics 18*(3), 457–490.

Schultz, P. W., J. M. Nolan, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological science 18*(5), 429–434.

Servátka, M. (2009). Separating reputation, social influence, and identification effects in a dictator game. *European Economic Review 53*(2), 197–209.

Sliwka, D. (2007). Trust as a signal of a social norm and the hidden costs of incentive schemes. *The American Economic Review 97*(3), 999–1012.

Van der Weele, J. (2009). The signaling power of sanctions in social dilemmas. *The Journal of Law, Economics, & Organization 28*(1), 103–126.

Vanberg, C. (2008). Why do people keep their promises? an experimental test of two explanations 1. *Econometrica 76*(6), 1467–1480.

Appendix A: Proof of Lemma 1

The first inequality follows from Proposition 2 in Milgrom (1981). Let us show the second inequality. Fix any $N' \in [0,1]$. We have

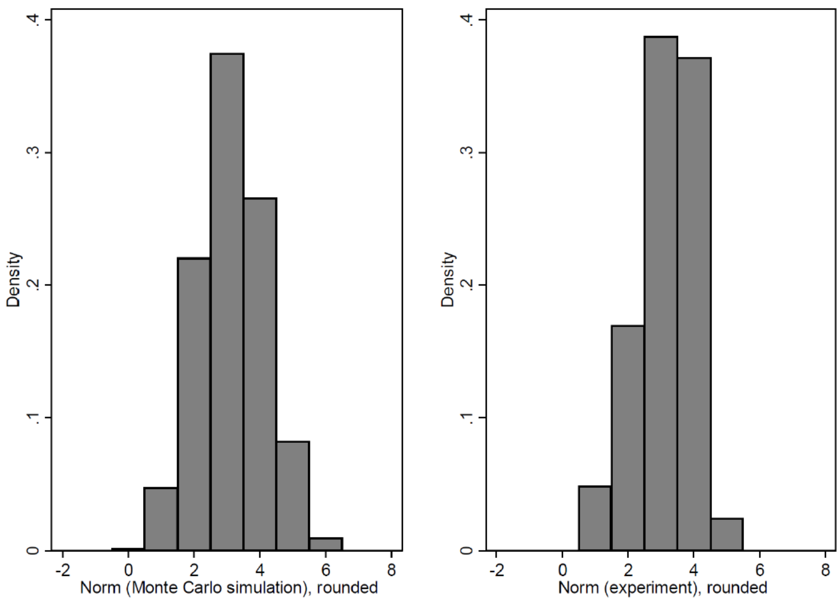
$$\begin{aligned} G(N'|s) &= \Pr[N \leq N'|s] \\ &= \frac{f[s|N \leq N']G(N')}{f[s|N \leq N']G(N') + f[s|N > N'](1-G(N'))} \\ &= \frac{1}{1+\frac{f[s|N>N'](1-G(N'))}{f[s|N\leq N']G(N')}} = \frac{1}{1+\frac{(1-G(N'))}{G(N')}\gamma(s)}, \end{aligned} \tag{12}$$

where the second equality is by Bayes rule, and $\gamma(s) \equiv \frac{f[s|N>N']}{f[s|N\leq N']}$. Then, by the law of total probability

$$\begin{aligned} \gamma(s) &= \frac{f[s|N > N']}{f[s|N \leq N']} = \frac{\int_{N'}^{1} f(s|x)g(x|x > N')dx}{\int_{0}^{N'} f(s|y)g(y|y \leq N')dy} \\ &= \frac{\frac{1}{1-G(N')}\int_{N'}^{1} f(s|x)g(x)dx}{\frac{1}{G(N')}\int_{0}^{N'} f(s|y)g(y)dy} \\ &= \frac{G(N')}{1-G(N')}\int_{N'}^{1} \frac{f(s|x)}{\int_{0}^{N'} f(s|y)g(y)dy} g(x)dx \\ &= \frac{G(N')}{1-G(N')}\int_{N'}^{1} \frac{1}{\int_{0}^{N'} \frac{f(s|y)}{f(s|x)} g(y)dy} g(x)dx. \end{aligned}$$

Since $x > y$ for any x and y under the integral signs (except for $x = y = N'$), (1) implies that $\frac{f(s|y)}{f(s|x)}$ strictly decreases with s. Consequently, the whole function $\gamma(s)$ strictly increases with s. Then, by (12) we obtain that $G(N'|s)$ strictly decreases with s. ■

Appendix B: Sampling distribution of average transfers in Experiment 1



Appendix C: Experimental instructions

C.1 Experiment 1

General information

Welcome to the experiment! The goal of this experiment is to study individual behavior in particular situations. If you have a question, please raise your hand. We will be glad to help you at your seat. **During the experiment, any other communication is not permitted!**

In this experiment, you can earn money. How much you earn depends on your decisions as well as on the decisions of the other participants. More detailed information about this is provided in the experimental instructions.

Your payoff will be paid to you personally in cash at the end of the experiment.

Your payoff and your decisions will be treated strictly confidentially. None of the participants will get to know during or after the experiment with whom he interacted. Your decisions are hence **anonymous**.

Experiment

This experiment consists of **three** parts.

Your payoff and the payoffs of the other participants are obtained from one of the three parts. This means that at the end of the experiment **one part will be randomly selected for all participants and paid out.**

Thus, thoroughly consider your decisions in each part of the experiment! Any of your decisions may result in a monetary payoff and therefore influence your today's income.

Next, you will receive instructions for the first and the second parts of the experiment. After the second part is over, you will receive instructions for the third part of the experiment.

Part 1 and Part 2

All participants are randomly divided into participants A and participants B. Every participant is matched to another person in the other role, so that **each participant A is matched to one participant B**. Both participants are seated in this room. The assignment of roles and the matching of participants to each other stays the same in part 1 and part 2. You will see on the computer screen which role you are assigned to.

Part 1

As described above, the earnings from part 1 will be paid out to all participants at the end of the experiment with probability 1/3 (otherwise, the earnings from part 2 or part 3 will be paid out).

In part 1, every participant receives a show-up fee of €2.50.

In addition to this, every participant receives an endowment of €5.

The task of all participants in part 1 is to guess the behavior of other participants *in part 2* of the experiment as precisely as possible. Every participant can earn an additional payoff by a good guess. You will find further information on this on your screen, after the rules for part 2 are explained.

The participants will be informed whether their guess has earned an additional payoff after the second part of the experiment.

Part 2

As described above, the earnings from part 2 will be paid out to all participants at the end of the experiment with probability 1/3 (otherwise, the earnings from part 1 or part 3 will be paid out).

In part 2, every participant receives a show-up fee of €2.50.

Decision of participant A:

Participant A receives an endowment of €14. He can give a part of his endowment to participant B.

Decision of participant B:

Participant B does not take any decision about the division of the endowment.

Therefore, the **payoffs** are calculated as follows:

Payoff to participant A = €14 − amount given to participant B

Payoff to participant B = Amount given by participant A

Participant B will be informed about the amount that was given to him by participant A only at the end of the experiment, namely after the third experimental part.

This is the end of the instructions for parts 1 and 2. Please take your time, and be sure to understand these instructions. **If you have any questions, raise your hand and an experimenter will come to you.**

[*Belief elicitation questions (shown on screen in part 1)*]

[*For participants B:*]

In what follows, you will be asked a question to provide a guess. If your guess to this question does not deviate from the true value by more than 15 cents, you will get a bonus of €5. The question refers to the behavior of the participants in this room. Please try to answer the question as best you can.

Question:

Please guess the **average** amount which will be given by participants A to participants B in part 2 of the experiment.

The average amount which participants A give to participants B is (from 0.00 to 14.00 Euro):

[*For participants A:*]

In what follows, you will be asked a question to provide a guess. If your guess to this question does not deviate from the true value by more than 15 cents, you will get a bonus of €5. The question refers to the behavior of the participants in this room. Please try to answer the question as best you can.

We have asked participants B the following question:

*"Please guess the **average** amount which will be given by participants A to participants B in part 2 of the experiment."*

Also participants B get a bonus of €5 for a guess which does not deviate from the true value by more than 15 cents.

Question:

Please guess the answer to this question of the participant B **who is matched to you** (namely, what do you think is the belief of the participant B who is matched to you about the average amount given by participants A?).

The amount which is expected by participant B is (from 0.00 to 14.00 Euro):

Part 3[21]

In the third part of the experiment, all participants retain their roles (participant A or participant B) which were previously assigned to them.

For every participant A, a random mechanism will select one participant B from this room who has not interacted with the participant A in the previous parts of the experiment. This person will be the recipient of the amount that participant A gives in part 3.

As described above, the earnings from part 3 will be paid out to all participants at the end of the experiment with probability 1/3 (otherwise, the earnings from part 1 or part 2 will be paid out).

In part 3, participant A receives an amount of €2.50.

At the beginning of part 3, participants B can decide whether or not their guess submitted in the first part (regarding the average amount which participants A send to participants B) may be transmitted to participant A. For the disclosure of this information participants B receive an amount of €2.50.

Decision of participant A:

As in part 2, participant A receives an endowment of €14. He can give a part of his endowment to the participant B who is now matched to him.

Decision of participant B:

Participant B does not take any decision about the division of the endowment.

Additionally, participant A receives the following information:

a) (If participant B has agreed to transmit his/her guess) The expectation of the currently matched participant B (namely, the recipient of the amount given in part 3)

[21] The paper instructions for part 3 were distributed after part 2 was over.

about the average amount which was given by participants A to participants B in part 2.

b) The average amount which was given by the other participants A in this room to participants B in part 2.

Therefore, the **payoffs** are calculated as follows:

Payoff to participant A = €14 − amount given to participant B

Payoff to participant B = Amount given by participant A

Participant B will be informed about the amount that was given to him by participant A at the end of the experiment.

This is the end of the instructions for the third part. Please take your time, and be sure to understand these instructions. **If you have any questions, raise your hand and an experimenter will come to you.**

C.2 Experiment 2

General information

Welcome to the experiment! The goal of this experiment is to study individual behavior in particular situations. If you have a question, please raise your hand. We will be glad to help you at your seat. **During the experiment, any other communication is not permitted!**

In this experiment, you can earn money. How much you earn depends on your decisions as well as on the decisions of the other participants. More detailed information about this is provided in the experimental instructions.

Your payoff will be paid to you personally in cash at the end of the experiment.

Your payoff and your decisions will be treated strictly confidentially. None of the participants will get to know during or after the experiment with whom he interacted. Your decisions are hence **anonymous**.

Experiment

This experiment consists of **three** parts.

Your payoff and the payoffs of the other participants are obtained from one of the three parts. This means that at the end of the experiment **one part will be randomly selected for all participants and paid out.** Herewith, you are guaranteed to get at least €4.

Thus, thoroughly consider your decisions in each part of the experiment! Any of your decisions may result in a monetary payoff and therefore influence your today's income.

Next, you will receive instructions for the first and the second parts of the experiment. After the second part is over, you will receive instructions for the third part of the experiment.

Part 1 and Part 2

All participants are randomly divided into participants A and participants B. Every participant is matched to another person in the other role, so that **each participant A is matched to one participant B**. Both participants are seated in this room. The assignment of roles and the matching of participants to each other stays the same in part 1 and part 2. You will see on the computer screen which role you are assigned to.

Part 1

As described above, the earnings from part 1 will be paid out to all participants at the end of the experiment with probability 1/3 (otherwise, the earnings from part 2 or part 3 will be paid out).

In part 1, every participant receives a show-up fee of €4.

In addition to this, every participant receives an endowment of €5.

The task of all participants in part 1 is to guess the behavior of other participants *in part 2* of the experiment as precisely as possible. Every participant can earn an additional payoff by a good guess. You will find further information on this on your screen, after the rules for part 2 are explained.

The participants will be informed whether their guess has earned an additional payoff after the second part of the experiment.

Part 2

As described above, the earnings from part 2 will be paid out to all participants at the end of the experiment with probability 1/3 (otherwise, the earnings from part 1 or part 3 will be paid out).

In part 2, every participant receives a show-up fee of €4.

Decision of participant A:

Participant A receives an endowment of €14. He can give a part of his endowment to participant B.

Decision of participant B:

Participant B does not take any decision about the division of the endowment.

Therefore, the **payoffs** are calculated as follows:

Payoff to participant A = €14 − amount given to participant B

Payoff to participant B = Amount given by participant A

Participant B will be informed about the amount that was given to him by participant A only at the end of the experiment, namely after the third experimental part.

This is the end of the instructions for parts 1 and 2. Please take your time, and be sure to understand these instructions. **If you have any questions, raise your hand and an experimenter will come to you.**

[*Belief elicitation questions (shown on screen in part 1)*]

[*For participants B:*]

In what follows, you will be asked a question to provide a guess. If your guess to this question does not deviate from the true value by more than 15 cents, you will get a bonus of €5. The question refers to the behavior of the participants in this room. Please try to answer the question as best you can.

Question:

Please guess the **average** amount which will be given by participants A to participants B in part 2 of the experiment.

The average amount which participants A give to participants B is (from 0.00 to 14.00 Euro):

[*For participants A:*]

In what follows, you will be asked a question to provide a guess. If your guess to this question does not deviate from the true value by more than 15 cents, you will get a bonus of €5. The question refers to the behavior of the participants in this room. Please try to answer the question as best you can.

We have asked participants B the following question:

*"Please guess the **average** amount which will be given by participants A to participants B in part 2 of the experiment."*

Also participants B get a bonus of €5 for a guess which does not deviate from the true value by more than 15 cents.

Question:

Please guess the answer to this question of the participant B **who is matched to you** (namely, what do you think is the belief of the participant B who is matched to you about the average amount given by participants A?).

The amount which is expected by participant B is (from 0.00 to 14.00 Euro):

Part 3[22]

In the third part of the experiment, all participants retain their roles (participant A or participant B) which were previously assigned to them.

For every participant A, a random mechanism will select one participant B from this room who has not interacted with the participant A in the previous parts of the experiment. This person will be the recipient of the amount that participant A gives in part 3.

As described above, the earnings from part 3 will be paid out to all participants at the end of the experiment with probability 1/3 (otherwise, the earnings from part 1 or part 2 will be paid out).

In part 3, participant A receives an amount of €4.

At the beginning of part 3, participants B can decide whether or not their guess submitted in the first part (regarding the average amount which participants A send to participants B) may be transmitted to participant A. For the disclosure of this information participants B receive an amount of €4.

Decision of participant A:

As in part 2, participant A receives an endowment of €14. He can give a part of his endowment to the participant B who is now matched to him.

Decision of participant B:

Participant B does not take any decision about the division of the endowment.

Additionally, participant A receives the following information about the expectations of participants B (if the corresponding participant B has agreed to transmit his/her guess):

a) The expectation of the currently matched participant B (namely, the recipient of the amount given in part 3) about the average amount which was given by participants A to participants B in part 2.

b) The expectation of another randomly selected participant B about the average amount which was given by participants A to participants B in part 2.

As in the case of the matched participant B, it is ruled out that the other randomly selected participant B has already interacted with the participant A in the previous parts of the experiment.

Therefore, the **payoffs** are calculated as follows:

[22] The paper instructions for part 3 were distributed after part 2 was over.

Payoff to participant A = €14 − amount given to participant B

Payoff to participant B = Amount given by participant A

Participant B, who is the recipient of the amount given by participant A, will be informed about the amount that was given to him by participant A at the end of the experiment.

This is the end of the instructions for the third part. Please take your time, and be sure to understand these instructions. **If you have any questions, raise your hand and an experimenter will come to you.**