Morell, Alexander

Working Paper

The short arm of guilt: Does it only hit who is close?

Preprints of the Max Planck Institute for Research on Collective Goods, No. 2014/19

Provided in Cooperation with:
Max Planck Institute for Research on Collective Goods

Suggested Citation: Morell, Alexander (2014) : The short arm of guilt: Does it only hit who is close?, Preprints of the Max Planck Institute for Research on Collective Goods, No. 2014/19, Max Planck Institute for Research on Collective Goods, Bonn

This Version is available at:
http://hdl.handle.net/10419/106906

Preprints of the
Max Planck Institute for
Research on Collective Goods
Bonn 2014/19

The Short Arm of Guilt: Does it only hit who is close?

Alexander Morell

MAX PLANCK SOCIETY

Preprints of the
Max Planck Institute
for Research on Collective Goods

Bonn 2014/19

The Short Arm of Guilt:
Does it only hit who is close?

Alexander Morell

November 2014

Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn
http://www.coll.mpg.de

The Short Arm of Guilt:
Does it only hit who is close?

Alexander Morell*

Abstract

In a laboratory experiment, I test whether guilt aversion, i.e., a preference to fulfill other people's expectations, plays out stronger if agents are socially close. I induce two different minimal group identities in participants and randomly assign participants to one of two treatments. Senders either play a dictator game with a receiver from their own group (ingroup treatment) or from the other group (outgroup treatment). I let senders condition their amount sent on second-order beliefs. I find that, in the realm of realistic beliefs (i.e., the sender expects the receiver to expect the sender to send no more than half of the pie), the positive influence of second-order beliefs on how much the sender sends is stronger in the ingroup treatment. In both treatments, about half of the senders remain unaffected by second-order beliefs. In the ingroup treatment, unaffected senders identify less with their group than affected senders do. This is not true for the outgroup treatment.

* Max Planck Institute for Research on Collective Goods, Bonn. Comments are welcome and can be sent to morell@coll.mpg.de. I thank Christoph Engel, Susann Fiedler, Sebastian Goerg, Oliver Kirchkamp, Anna Kochanova, Michael Kurschilgen, and Franziska Tausch, as well as the participants of a workshop at the University of Cologne for helpful discussions and comments on earlier versions of the paper.

Motivation

A couple of years ago, the motivational power of people's desire to avoid guilt reached the economic literature and was called guilt aversion. Among other things, guilt aversion was used to explain why people keep their promises in an investment game (trust game) and it was modeled as a preference to choose actions conforming with another person's expectation (Battigalli & Dufwenberg, 2007; Charness & Dufwenberg, 2006). Surprisingly, the economic model of guilt aversion has not granted a lot of attention to the relevance of social distance. Why should I feel guilty with regard to someone I do not care about? Or why should I feel guilty about letting someone down if I was not responsible for his well-being? Psychologists treat the sensitivity of people to others' expectations as inherently context-dependent. And this context is shaped by relationships. In fact, Charness and Dufwenberg (2006) motivate their research on guilt by citing articles from psychology which deals with the impact of relationships on guilt and postulates a "communal relationship" as precondition to experiencing guilt vis-à-vis a person (as opposed to a mere exchange relationship), for instance Baumeister, Stillwell, & Heatherton (1994). Other psychological theories postulate the expansion of one's own self to that of a relationship partner as a prerequisite to experience guilt (Aron, Aron, & Norman, 2003; Aron, Aron, Tudor, & Nelson, 1991).

I assess the impact of social closeness on guilt aversion in a dictator game by systematically varying shared social identity using a minimal group paradigm. If even group identity induced in the lab changes how second-order beliefs (i.e., the sender's expectations over the receiver's expectations) induce actions, the effect is likely to be stronger by magnitudes in the field.

In each session, I allocate half of the participants to one and the other half to a second group according to their respective stated preference for one of two modern painters. I further reinforce their respective group identity (Tajfel & Turner, 1979) by letting the two groups compete for a prize in a real-effort task and subsequently measure to what extent participants identify with their group. Then I randomly match participants in pairs, half of them with a partner from their own group (ingroup treatment) and the other half with a partner from the other group (outgroup treatment). These pairs play the following variant of a dictator game. From receivers I elicit their belief about what they expect the sender to send to a receiver in their situation (from the same group or from the other group, respectively). Then, by means of the strategy method, I elicit what share of a pie of 100 "Talers" a sender wishes to send to the receiver conditional on 11 possible second-order beliefs. I look at realistic second-order beliefs in isolation first. I speak of realistic second-order beliefs if the sender expects the receiver to expect the sender to send half of the pie at most because amounts sent that exceed half of the sender's endowment are extremely rare (Engel, 2011). In the realm of realistic beliefs, I find that the influence of second-order beliefs on the amount sent in a dictator game is stronger if the receiver shares the sender's group identity. For the analysis of exaggerated second-order beliefs, the reader is referred to Appendix A. I speak of exaggerated second-order beliefs if the sender expects the receiver to expect the sender to send more than half of the pie.

In both treatments, about half of the senders remain unaffected by second-order beliefs. If senders and receivers are from the same group, unaffected senders are characterized by lower degrees of group identification. This is not true for the treatment where senders and receivers are from different groups.

My findings are relevant for certain questions that have been studied for a long time, yet remain unresolved. These findings concern promises, ingroup favoritism and the theory of guilt aversion.

The results of this experiment suggest that people may keep unenforceable promises because the promise establishes a closer relationship between the parties. My amendment of guilt aversion reconciles the contradicting results of Charness & Dufwenberg (2006; 2010), on the one hand, and those of Vanberg (2008), on the other. Charness & Dufenberg (2006; 2010) explained why people keep a promise to a stranger by the fact that promisors know that the promise raises the promissee's expectations. Vanberg (2008) found that promisors keep their own promise, but did not respond to second-order expectations induced by a thirdparty's promise. If guilt aversion depends on social integration, Vanberg's result can be explained by a lack of social integration between a promisor with a third-party promissee. That lack of social integration in turn leads to a lack of influence of second order beliefs on action.

Güth, Ploner & Regner as well as Ockenfels & Werner, theorise that participants favour fellow group members because senders know that recievers expect more in ingroup interactions than in outgroup interactions. According to my findings, this explanation is at least incomplete. Explanations along the lines of Güth et al. and Ockenfels & Werner should take into account that second-order beliefs also matter more in ingroup interactions than they do in outgroup interactions. If second-order beliefs have a stronger effect ingroup than outgroup one would predicts ingroup favoritism to arise even if receiver expectations are held constant. The only prerequisite would be that group identity is strong enough.

My main finding that shared identity induces the influence of second-order beliefs on action explains why guilt aversion has been rejected in anonymous experiments (i.e., Ellingsen, Johannesson, Tjøtta, & Torsvik, 2010), while it has been confirmed in experiments that allow for some form of relationship between participants (Charness & Dufwenberg 2006, Reuben et al. 2009). This finding structures the hitherto inconclusive literature on guilt aversion (for a review, see below). My finding also moderates conclusions that guilt aversion has been rejected in its entirety and specifies the realm of application of guilt aversion to social interaction across a small social distance.

Literature

This paper will investigate whether a model of guilt aversion, which is sensitive to the degree of social intergration between agents, leads to better predictions than the existing formulations

of the theory. The blind spot of guilt aversion with regard to relationships seems to have led to inappropriate test beds when testing the theory. The beginnings of the lab career of guilt aversion were promising (Charness & Dufwenberg, 2006). However, further tests of the theory of guilt aversion in the lab led to mixed results: Vanberg (2008) conducts a dictator game experiment where he claims to separate the effect of the mere promise and that of second-order beliefs. He reshuffles half of the sender-receiver pairs after the promise, leaving only the receivers uninformed about whether their pair has been reshuffled. This leaves receivers' beliefs constant across treatments. Treatments only vary by whether the sender is bound by a promise to the receiver or not. Vanberg finds an effect of the promise, although second-order beliefs are constant over treatments. Although in his appendix he presents some evidence that second-order beliefs correlate with action, he cannot show a causal effect of second-order beliefs on action. He concludes that second-order beliefs cannot explain the effect of promises in trust games. Reuben et al. (2009) conduct an experiment in which they elicit investors' beliefs in a trust game and report them to trustees. They do find evidence in favor of guilt aversion. Ellingsen et al. (2010) conduct a series of dictator-game and investment-game experiments where receivers report their beliefs on the amount sent to the experimenter and the experimenter reports these beliefs to the senders, inducing second-order beliefs. Ellingsen and coauthors find evidence that second-order beliefs do not determine action, but that actions induce second-order beliefs. Lately even prominent promoters of guilt aversion, Charness and Dufwenberg (2010), merely found "limited support" for guilt aversion. And finally, in a trust game with an investor, a trustee, and two inactive players, Bellemare et al. (2011) found trustees to have a positive willingness to pay to avoid guilt vis-à-vis the investor only. At first glance, the literature could lead the reader to believe that the correlation between second-order beliefs and actions is an instable phenomenon that tends to be revealed as a confound – either with a preference to keep a promise (Vanberg, 2008) or with a (false) consensus effect (Ellingsen et al., 2010).

However, re-analyzing the experiments just mentioned with regard to the intensity of the relationship between subjects and the findings of guilt aversion, a correlation seems to emerge. Charness and Dufwenberg (2006) use pre-play communication by means of a one-page free text letter in a classroom experiment where subjects can see each other. This protocol is apt to make participants feel closer to each other. They find second-order beliefs to correlate with actions. Vanberg (2008) claims to disentangle the effect of the promise from that of expectations by rematching half of the participants randomly after communication in an anonymous, computerized dictator game. However, this protocol does not merely destroy the promise of randomly rematched dictators; rather, it also destroys the social relationship participants may have built through the promise. Vanberg concludes that people have a preference for keeping a promise, independently of second-order beliefs. Reuben et al. (2009) use 56 subjects in one session, all of whom were MBA Students at the Kellogg Business School. This school does not have more than 650 students in total. Given that MBA programs are meant to establish close networks among their students, it is not unlikely that there was some esprit de corps connecting the subjects in this setting. Accordingly, Reuben et al. (2009) find evidence of

guilt aversion. Ellingsen et al. (2010) use an anonymous double blind protocol. The only contact amongst participants is that beliefs elicited from the receiver are reported to senders – a procedure specifically meant to exclude any social integration of participants. Reuben et al. do not find any correlation between second-order beliefs and action. Charness and Dufwenberg (2010) adopt a protocol enabling senders either to make a promise to receivers by sending a pre-formulated sheet of paper or not to make a promise by sending an empty sheet of paper. This procedure does not allow for any personalized contact between participants, but involves a promise. The anonymity of this procedure may be the reason why Charness and Dufwenberg (2010) do not find clear support for guilt aversion in their experiment. Bellemare et al. (2011) only find guilt aversion of the trustees vis-à-vis the investors. But the trustees have no willingness to pay to avoid guilt vis-à-vis the inactive players. In fact, although the setup of the experiment is anonymous in that subjects participated online from their homes, the contrast between the investor who actually does act with effect on the trustee and the inactive players who does not may have induced the trustees to feel closer to the investors.

Of course, this juxtaposition of six experiments is far from being conclusive evidence for guilt aversion only to play out if agents are socially integrated. But it may be a hint. And given that the psychological theories from which the economic theory of guilt aversion was originally derived accord a prominent role to relationship, the hint merits to be taken seriously. A serious test of this hint seems all the more warranted as some parts of the literature fit the pattern found in the six cited papers less well. Dufwenberg and Gneezy (2000) find that the trustees' second-order beliefs correlate with actions in a trust game, although they apply a double blind and thus very anonymous procedure.1

There is a large literature showing that decreasing social distance (Charness & Gneezy, 2008; Frey & Bohnet, 1999; Hoffman, McCabe, & Smith, 1996; Leider, Möbius, Rosenblat, & Do, 2009, 2010; Rankin, 2006), increasing social integration (Brañas-Garza et al., 2010), or inducing a common group identity (Chen & Li, 2009; Dawes, Van de Kragt, & Orbell, 1988) between participants leads social preferences to play out more strongly in dictator games. But none of the cited studies treats belief-dependent preferences.

The literature studying the relevance of second-order beliefs has not produced an answer either to the question whether guilt aversion requires some form of social closeness. Rankin (2006) studies whether receivers' demands in a dictator game have different effects if receivers and senders communicate face to face or anonymously. I study the effect of second-order beliefs and not of demands. In contrast to outright demands, the beliefs I work with do not have any normative or imperative appeal. Bicchieri and Chavez (2010) study the impact of first- and second-order beliefs on transfers in an ultimatum game. In the ultimatum game, however, second-order beliefs are strategically relevant. Guilt aversion claims an influence of strategically irrelevant beliefs. Therefore the study does not provide evidence on whether

1 They neither control for the false consensus effect, nor can they claim causality of second-order beliefs. So causality could also run from action to beliefs or there could be a confounding variable.

people have a preference to act in accordance with second-order beliefs. Recently the idea arose that ingroup favoritism is caused by changes in second-order beliefs. Güth et al. (2009) as well as Ockenfels and Werner (in press) hypothesize that senders treat ingroup receivers preferentially because they know that ingroup receivers expect them to send more. Güth et al. do not find clear support for this hypothesis. Ockenfels and Werner find that indeed senders treat ingroup receivers better if the latter know that they share the sender's group identity. This effect is attenuated if the receiver does not know of the sender's group identity. I am not interested in the effect of changing levels of second-order beliefs on action. I ask for the effect of shared group identity on the capacity of second-order beliefs to induce action.

Experiment

I run an experiment with two treatments, an ingroup treatment and an outgroup treatment. The on-screen experiment is programmed in z-tree (Fischbacher, 2007). Group membership is induced, conditional on the subjects' preferences for paintings (Chen & Li, 2009), and reinforced by letting groups compete in a real-effort task (Rockenbach, Böhm, & Weiss, 2013). Each of the 15 sessions conducted in the Bonn EconLab comprises 16 participants recruited via ORSEE (Greiner, 2004). Each subject receives a show-up fee of 4 €. In the experiment, the subjects play for the experimental currency "Talers". Participants are paid in Euros. 1 Taler converts to 0.11 €. The experiment proceeds as follows.

All participants are seated in front of a computer terminal, separated by cubicles. They are first asked to compare paintings by Klee and Kandinsky (group segregation stage); then, they compete in a real-effort task (group reinforcement stage) and fill out a pre-experiment questionnaire (questionnaire stage); and finally, they play a dictator game (game stage).

Group Segregation Stage

In the group segregation stage, each participant is assigned to one of two groups according to his/her preference for one of two painters – Klee and Kandinsky. The procedure is adapted from Chen and Li (2009): For five pairs of paintings by Klee and Kandinsky (the same paintings as in Chen and Li (2009)), subjects are asked to state how much they prefer one painting to another. To answer this question, participants use a slider bar (labeled in three steps [L=left, R=right]: I strongly prefer L, I like both paintings equally, I strongly prefer R). The position of the slider bar is translated into a distribution of 10 points between the two paintings (I strongly prefer L = 10 points to L, I like both paintings equally = 5 to L and 5 to R, etc.). Then, for each participant, the points allocated to Klee are summed up. The same is done for the points allocated to Kandinsky. Subsequently the computer labels that half of participants who allocated the highest amounts of points to Klee the "Klee group" and labels the

other half the "Kandinsky group". Participants then are informed about their group membership.[2]

Group Reinforcement Stage

In the group reinforcement stage, the two groups compete against each other in a real-effort task to intensify the perception of belonging to a group by experiencing interdependence and a common fate. The task subjects compete in is the following (Rockenbach et al., 2013). Participants receive a 15-page text. Then, on their screens, I ask them for letters in the text that I define by page, line, word, and position. Participants have four minutes to identify as many letters as they can. The group that jointly accumulates the larger number of correct answers wins. Each participant of the winning group receives 26 Talers. If the groups tie, all participants receive 13 Talers. Participants do not receive any feedback on the between-group competition until the very end of the experiment, which is why independence of observations is preserverd.

Questionnaire Stage

The computer randomly assigns half of each group to the role A (sender) and the other half of each group to the role B (receiver). Then the computer randomly pairs each sender with a receiver. Half of the senders will be paired with a receiver from their own group (ingroup treatment) and half of the senders will be paired with a receiver from the other group (outgroup treatment).

In the questionnaire stage, the senders (A) are not informed about their role. They answer a questionnaire on how much they identify with their group (Doosje et al. 1995; see appendix B.3.2b. for questions).

The receivers (B) are informed about their role and the group membership of the participant in role A they have been paired with. On their screens they are informed that in the subsequent stage they will be paired with a sender, which group this sender belongs to, and that the sender can freely split 100 Talers between himself and the receiver. They are then asked to predict the average amount a person in their situation – i.e., a receiver paired with a sender from the same [the other] group – would receive in the experiment. The responder can enter any guess, which can be expressed in a full amount between 0 and 100 Talers. It is announced that each subject who predicts an average amount (which is no more than 1 Taler off the actual average

2 Ties were resolved by the order of the randomly assigned subject ID. Subjects could allocate fractions of points, so that virtually an infinite number of possible sums of points for Klee paintings were possible. The slider bar did not have a scale beyond the three labels mentioned in the text. Therefore it was virtually impossible to set it to precise integers (other than 10.0; 5.5; 0.10). Because ties thus were extremely unlikely, the tie-breaking rule was not included in the instructions. However, a tie occurred twice.
Of 120 senders, ten ended up in the Kandinsky group although they had awarded more points to Klee. Two who had given exactly the same amount to Klee as they gave to Kandinsky ended up in the Kandinsky group.

amount received by receivers in their situation during the session) will receive an extra payment of 125 Talers = 13.75 € (see Appendix B.3.2.a for details). After all participants have completed their respective questionnaires, the questionnaire stage ends.

Game Stage

Subjects play a sender-receiver game (dictator game). On the first screen, all participants are informed about their role. Also, all participants are reminded about their own group membership and informed about the group membership of the participant they have been paired with. Senders receive a pie of 100 Talers. They can send any share to their respective receiver, which can be expressed in full Talers. The amount sent is elicited by means of a strategy method. Senders are asked what they would like to send, conditional on their respective receiver's belief. They express the amount they wish to send for the receiver's beliefs of 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 Talers. It is explained to them that the computer will activate the choice closest to the receiver's actual stated belief.3 After the senders have filled in the strategy vector, a screen will reveal the true stated belief of the sender, and the computer will put into effect the allocation for the case closest to that belief.

The use of the strategy method described does not confound the treatment effect with an experimenter demand effect, because the experimenter's "demand" to condition the amounts sent on second-order beliefs remains constant over treatments.

Apart from my treatment manipulation (ingroup vs. outgoup) and computerization, the design described amends that of the paper that mainly motivated this work (Ellingsen et al., 2010) merely in that I use the strategy method to let senders condition their amounts sent on different second-order beliefs instead of just one. Generally, this method of letting senders condition their amounts sent on receivers' previously stated beliefs seems to be standard in the literature on guilt aversion (Bellemare, Sebald, & Suetens, 2013; Ellingsen et al., 2010; Reuben et al., 2009).

Finally, in a posttest, I measured perceived closeness between the sender and the receiver, using a one-item test by Aron et al. (2003), and I ask participants for some demographic data, such as gender, age, and occupation.

Theory and Hypothesis

To illustrate my theoretical point, I use the simplest formulation of guilt aversion. It can be found in Charness and Dufwenberg (2006). I will extend their model to include social distance as a driving force of guilt aversion. I will justify my amendment with the help of psychological and economic theory. Finally I will derive and specify the hypothesis of my exper-

3 Receivers' stated beliefs are rounded to 0 or the closest multiple of 10 according to general rounding conventions.

iment. For conjectures about what could plausibly be expected to happen in the realm of exaggerated beliefs, see Appendix A.

Theory

For a simple two-strategy dictator game, Charness and Dufwenberg (2006) propose a simple definition of the utility uS of the outcome given the sender S choses A (the selfish option) over B (the generous option):

$$uS = \pi A - \gamma S \cdot \pi B \cdot \tau S$$

The sender's utility is increased by his money payoff of choice A, but it is reduced by a guilt term. The sender's money payoff of option B is denoted by πB,. The sender's sensitivity to guilt is $\gamma S \in [0,1]$. Finally, $\tau S \in [0,1]$ denotes the sender's second-order belief about how likely he thinks the receiver believes the sender to choose the generous option. The guilt term increases in all of these variables.

This utility function claims second-order beliefs influence the attractiveness of the selfish choice. Senders therefore should always respond to information from which they derive their second-order beliefs by being more or less prone to act more selfishly. We have not observed this consistently in experiments. Ellingsen et al. (2010), in particular, conducted a high-powered experiment that could not show any reactions of senders to second-order beliefs. The hunch derived from the literature above was that possibly people only condition their action on second-order beliefs if they interact with somebody to whom they feel close. Accordingly, I propose to let the guilt term in the utility function also depend on a measure of social closeness. Multiplying the guilt aversion part of the utility with $\alpha S \geq 0$, where αS is the sender's appreciation of how close his relationship with the receiver is, would express this dependence. The larger αS, the closer the relationship. An αS that is equal to 1 expresses guilt aversion as Charness and Dufwenberg (2006) defined it. αS equal to 0 would indicate a social distance too large to trigger any feeling of guilt.

$$uS = \pi A - \gamma S \cdot \pi B \cdot \tau S \cdot \alpha S$$

The justification of this extension lies in a recombination of the psychological theory of how closeness translates into empathy with the theory of reference-dependent preferences. In economics, other-regarding preferences à la Charness and Rabin (2002) or Fehr and Schmidt (1999) have been modeled to include the other's payoffs into the self's utility function. Guilt aversion goes a step further including the self's beliefs about the other's expectations into the utility function. Why do beliefs about expectations matter at all? Guilt aversion does not take a clear position on this question. One answer is that beliefs are important because we know the other's expectations matter for the other's utility. According to reference-dependent utility, expectations shift reference points (Abeler, Falk, Götte, & Huffman, 2009; Köszegi & Rabin, 2006). And outcomes below the reference point are coded as losses, while those above

are coded as gains. Losses loom larger than gains so that any outcome short of the expectation would have a strong negative impact on utility. Writing the impact of expectations on the other's utility into the self's utility would mean including second-order expectations in the self's utility function. In fact, the self's utility would include parts of the other's (reference-dependent) utility (instead of just plugging the other's payoff into the self's utility function). Economists seem reluctant to integrate the other's utility (as opposed to the other's payoffs) into the self's utility function because this would yield complex interdependence of agents' utility, rendering these utility functions difficult to use. Psychology has been bolder and has developed theories that are equivalent to the self integrating the other's utility into her utility functions. Psychologists frame the integration of the other's utility function into the self's as "self expansion", meaning the extension of one's own self to encompass other individuals' selves (Aron et al., 2003, 1991; Hewstone, Stroebe, & Jonas, 2008). In particular "participants in a close relationship include each other into their psychological selves" (Aron et al., 2003). Other authors describe that same thing, saying that "oneness" increases among ingroup members (Brewer, 2007). In a slightly different approach, focused on norms rather than utility, Baumeister et al. (1994) claim that guilt only arises due to a violation of norms induced by a "communal relationship". "Communal relationships are defined by the existence of implicit rules that the individuals must be concerned about each other's welfare (…). As a result, communal relationship partners do things simply to benefit each other without expecting equal or immediate benefits in return" (Baumeister et al., 1994). Psychological theory predicts that with a sufficiently close relationship between the self and the other comes the self's concern for the other's welfare. Accordingly, in my experiment, senders of the ingroup treatment would be expected to experience the utility they cause in receivers of their own group as their own utility to some degree. If the receivers' utility depends on their expectations, as reference-dependent utility suggests, this means that, according to self-extension theory, senders should behave in line with receivers' expectations to a larger extent if receivers are from their own group than if they are from a different group.

Hypotheses

H1: In the range of reasonable beliefs, second-order beliefs influence actions positively.

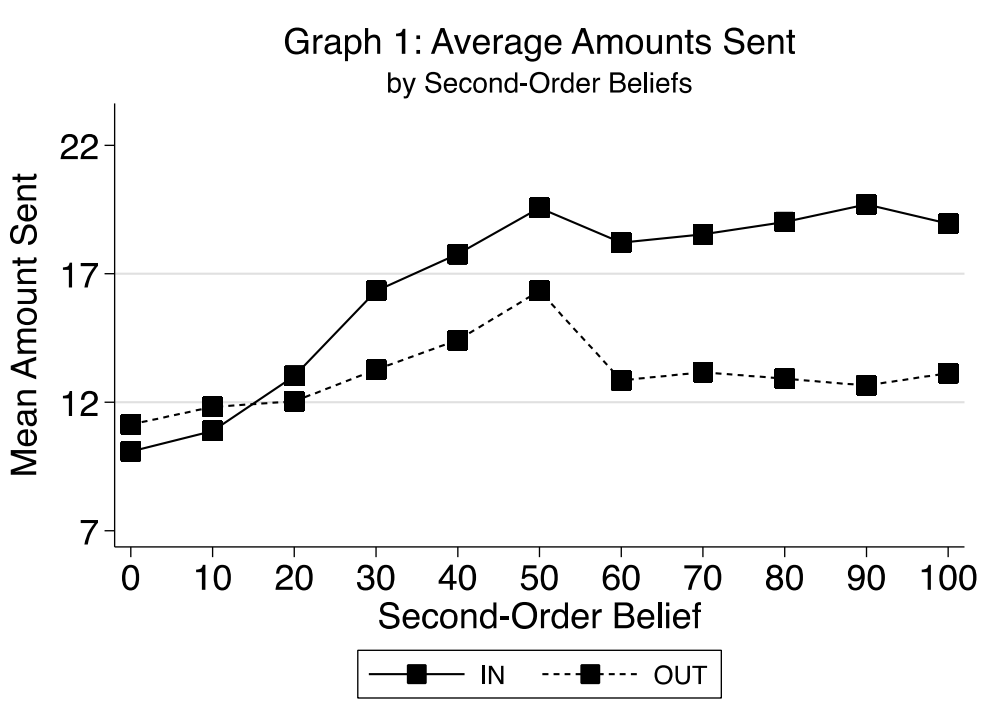
H2: This influence is stronger if the sender and the receiver are from the same group.

In the dictator game played, the range of reasonable beliefs goes from the sender sending 0% of the pie to the sender sending 50% of the pie on average. Shares sent that exceed this range are extremely rare (Engel, 2011) and, accordingly, it is extremely unlikely for receivers to expect the sender to send more than half of the pie.

Results

Of the 120 senders, 61 were female and 59 male. 114 senders were students, while 6 were not. The mean age was 24.15, the median age was 23, and the standard deviation 5.18 years.

After briefy exposing my analysis strategy, I test H1 and H2. For the exploration of the interaction effect of second-order beliefs and group identity on action in the realm of exaggerated beliefs, the reader is referred to Appendix A. At the end of this section, I analyze some additional findings on ingroup favoritism, the persistence of group identity, and participants unaffected by second-order beliefs.



Analysis Strategy

I analyze the decisions that 120 senders took for eleven possible second-order beliefs. 60 senders took their decisions in the outgroup treatment and 60 in the ingroup treatment.

I expect a main effect of the ingroup treatment, which is due to the famous phenomenon of ingroup favoritism (Tajfel, 1979). This effect has to be separated thoroughly from the interaction effect of the ingroup treatment with second-order beliefs, which tests my hypothesis. Therefore I will estimate a linear regression model of the following form with random effects on the participant level. The regression predicts amounts sent using second-order beliefs (sobelief), a dummy for the treatments (ingroup), and the interaction of both (sobelief*ingroup) as independent variables. The ingroup dummy is equal to 1 if the sender and the receiver are from the same group and 0 otherwise. I use a random-effects estimator instead of a fixed-effects estimator, insofar as it does not bias the results away from the result of the fixed-effects estimator. Using a fixed-effects estimator, one of my two main effects (ingroup) would

drop out because of a lack of within-subject variance. Therefore the purpose of using random effects is merely to make all effects visible in one model.

Amount Sent = β0 + β1*sobelief + β2*ingroup + β3*sobelief*ingroup + error (participant random effects) + error (residuals)

The coefficient β1 should pick up any effect of second-order beliefs which does not depend on the treatment manipulations. A positive and significant β1 in the real of realistic beliefs is evidence in favor of H1. β2 should pick up the level effect induced by in-group favoritism. And finally, β3 should pick up whether the ingroup vs. outgroup manipulation reinforces and attenuates the effect of second-order beliefs on action. If β3 is positive and significant in the realm of realistic beliefs, this is evidence in favor of H2 this paper set out to test.

On top of the simple linear regression model, I will report results from a Tobit regression with random effects at the participant level. It has been shown in the literature that if the experimental design allows not only sending to receivers, but also taking from them, some senders do actually "steal" from receivers' endowments (Engel, 2011). In my experiment, senders cannot transfer negative amounts. Therefore the data may be censored at zero. Tobit is a common approach to take this censoring into account.

For all random-effects linear regression models, the Hausman test is insignificant. In fact, all coefficients in the linear random effects model are virtually identical to those of the linear fixed effects model.

Testing the Predictions of Guilt Aversion over Reasonable Beliefs

Graph 1 shows that, in the realm of realistic second-order beliefs (0-50%), the latter influence the amount sent positively in both the ingroup (IN) and the outgroup (OUT) treatment. It is also clearly visible that this influence is stronger ingroup than it is outgroup. Both results are confirmed by the regression analysis summed up in table 1.

Table 1
Data for second-order beliefs from 0 to 50%

Dependent variable: **Amount sent**	Model 1: Linear regression	Model 2: Tobit regression
Second-order belief	0.100*** (0.0266)	0.141** (0.0431)
Ingroup (dummy)	-1.155 (3.039)	-5.756 (6.172)
Ingroup x Second-order belief	0.104** (0.0377)	0.207*** (0.0623)
Random effects at participant level	Yes	Yes
Constant	10.66***	-2.321
N	720	720
Number of Groups	120	120

*=p<0.05; **=p<0.01; ***=p<0.001
Standard errors in parenthesis

The number of total observations is 720. "Number of groups" refers to the observations grouped by individual participants. The regressions consider amounts sent for six different second-order beliefs per sender (0, 10, 20, 30, 40, 50).

The null hypothesis that in the realm of realistic beliefs second-order beliefs do not influence actions positively has to be rejected. But I am reluctant to interpret this result as general evidence in favor of guilt aversion. By using the strategy method, I basically asked senders to condition their amount sent on second-order beliefs, so it is hard to be surprised that they did.

More importantly, the null that the influence of second-order beliefs is not stronger if the sender and the receiver are from the same group has to be rejected. The "demand" to condition amounts sent on second-order beliefs was constant over treatments, so the positive and significant interaction effect of second-order beliefs and group identity remains valid evidence for the influence of shared identity on guilt aversion. Accordingly, I derive results one and two.

Result 1: In the realm of realistic beliefs the effect of second-order beliefs on action is positive. (Random effects regression, $p<0.001$, $\beta1=0.1$).

Result 2: This effect is stronger if senders and receivers share a common group identity. (Random effects regression, $p<0.01$, $\beta3=0.103$).

My results can also be shown by a Tobit random-effects regression which accounts for the possibility that senders would actually have taken money from receivers if I had let them (Model 2, *tobit random effects: $\beta1=0.14$, $p<0.01$; $\beta3=0.2$, $p<0.001$*).

Ex post, the fact that 97.5% of receivers stated beliefs below or equal to 50% of the pie can be regarded as a justification for generating hypotheses only for the realm of realistic beliefs (mode of elicited receiver beliefs: 0% in both treatments; mean overall: 22.30; mean ingroup:

25.35; mean outgroup: 19.26. The difference is marginally significant: Wilcoxon ranksum, N=120, p=0.073).

Additional Results on Ingroup Favoritism, Unaffected Participants, and the Persistence of Group Identity

It seems striking that the expected level effect of shared group identity is neither visible in the graph nor in the regression analysis. Even searching for differences between the levels of amounts sent by treatment and by second-order beliefs does not provide any statistically significant results (ranksum, all p>.22). For very low second-order beliefs (0% and 10%), senders even tend to send less on average to ingroup receivers than to outgroup receivers, although this difference is not statistically significant. Accordingly, the treatment coefficient in the regression is negative.

Result 3 (Null result): The results are inconsistent with general, i.e., belief-independent, ingroup favoritism. For no single level of second-order beliefs do ingroup senders send significantly more than outgroup sender (ranksum, all p>.22).

The results also show that in both treatments slightly less than half of the senders are completely unaffected by second-order beliefs. I call a sender "unaffected" if she/he intends to send the same amount for all eleven second-order beliefs offered in the strategy method. Of these participants, there are 29 of 60 in the ingroup treatment and 26 of 60 in the outgroup treatment. One may have expected the amount of unaffected participants to be smaller in the ingroup treatment than in the outgroup treatment because the theory says that second-order beliefs have more effect on action ingroup. But as the utility function set out above also contains a parameter for individual sensitivity to guilt, the slightly higher amount of unaffected participants in the ingroup treatment could easily be explained by a slightly greater number of insensitive participants who were randomly allocated to the ingroup treatment. In any event, the difference between treatments in unaffected participants is very small and statistically insignificant *(Chi-squared test, Chi2=0.3021, p>.58)*.

I also find that in the ingroup treatment unaffected participants identify less with their group than the affected senders *(Wilcoxon ranksum, p<.04)*. I measured the degree of group identification as the average score in the four questions of group identification presented to senders in the questionnaire stage (appendix B.3.2b.) In the outgroup treatment, unaffected participants do not distinguish themselves from the affected participants by the degree of group identification *(Wilcoxon ranksum, p>.55)*. The finding that affectedness and the degree of group identification correlate (only) in the ingroup treatment is in line with the theory set out above: In the outgroup treatment, the identification with one's fellow group members is irrelevant as the senders do not interact with their fellow group members. So only in the ingroup treatment should the degree of identification matter for what senders do. The stronger the identification with one's group in the ingroup treatment, the more second-order beliefs should determine action and the less likely it is that a sender does not react to second-order beliefs at all.

Result 4: In the ingroup treatment, unaffected senders (i.e., senders sending the same amount irrespective of the second-order belief) identify less with their group than affected senders (Wilcoxon ranksum, N=60, p<0.04). This is not the case in the outgroup treatment (Wilcoxon ranksum, N=60, p>.55).

A fifth and rather unexpected finding is that the senders' feeling of shared group identity seems to have vanished by the time they have completed their decisions in the game stage. In the posttest, I do not find any treatment difference between the senders' perceived closeness towards their respective receivers. Indeed, the outgroup senders seem to feel slightly closer to their receivers (mean score: 2.56) than ingroup senders seem to do (mean score: 2.53). Given that the treatment manipulation led to normal levels of group identification and does induce a significant difference between ingroup and outgroup treatment, the failure of the closeness measure to pick up a difference between the treatment groups ex post may merely mean that it is not a very reliable measure of group identity. However, it may also indicate that shared minimal group identity decays very quickly, stressing that I chose a very gentle intervention.

Discussion and Conclusion

To the best of my knowledge, this experiment is the first to show that, in the realm of realistic beliefs, social closeness – implemented here as shared group identity – determines how strongly senders' second-order beliefs influence the amounts sent in a dictator game. I used a minimal group paradigm to induce a shared identity and reinforced it slightly. The total intervention is extremely faint. Therefore, the effect is likely to be a lot stronger in the field, where relationships are based on family ties, friendship, co-workership, and the like. Also being class mates in an MBA program (Reuben et al., 2009), exchanging a one page letter (Charness & Dufwenberg, 2006) or being parties to a promise arising in a computer chat (Vanberg, 2008) are protocols that are likely to induce stronger shared identity than my treatment manipulation.

I further find that if senders and receivers are from the same group, those senders who previously stated that they identify strongly with the group are more likely to be affected by second-order beliefs. This further corroborates my main result that shared identity determines the effect of second-order beliefs on action.

My results clarify that guilt aversion will make better predictions in contexts of social closeness (families, friendships, co-workers) than in anonymous contexts (anonymous market transactions). On the one hand, they reveal that experiments in a very anonymous setting may be the wrong test bed to test theories of second-order belief-dependent preferences. On the other hand, my results suggest that theories on guilt aversion should spell out that social closeness is crucial for the effect of second-order beliefs on action.

In my experiment, a general level effect of ingroup favoritism is absent. Holding second-order beliefs fixed, I cannot find ingroup favoritism for any single level of second-order beliefs. This is in line with what Güth et al. (2009) and Ockenfels & Werner (in press) have suggested: Ingroup favoritism possibly does depend on second-order beliefs, such that the difference in amounts sent ingroup and outgroup are due to a higher level of second-order beliefs in ingroup interactions. Also in line with Güth et al. (2009) and Ockenfels & Werner (in press), I show that shared group identity does indeed translate into elevated expectations of receivers – which senders may well anticipate. But beyond Güth et al. (2009) and Ockenfels & Werner (in press), I also show in this experiment that their explanation may at least be incomplete. I show that shared group identity not only raises receivers' expectations, but leads to a stronger influence of the senders' second-order belief actions. Both effects together may just reinforce each other. However, according to my results, the increase in expectations required to trigger ingroup favoritism may be smaller than implied by Güth et al. (2009) and Ockenfels et al. (in press). In fact, my results suggest that ingroup favoritism independent of second-order belief remains possible in case of a strong shared identity. In case of very strong shared identity, ingroup favoritism could arise only through the stronger impact of a fixed level of second-order beliefs. This would mean that ingroup favoritism was possible under identical second-order beliefs ingroup and outgroup. Güth et al. (2009) and Ockenfels & Werner (in press) would not make this prediction. My experiment, which only induced group identity very gently, could not test this prediction conclusively. But this test appears to be a promising avenue for future research.

From my results it appears plausible that people hold a promise to a stranger because the promise creates a shared identity between the two, causing second-order beliefs to induce action. In future research it should be tested whether promise keeping can be better predicted by a theory of guilt aversion amended along the lines described here or by a preference to hold a promise. Promises that activate guilt aversion by creating a relationship between the parties would be compatible with a theory of "lexicographic promise keeping" proposed by Ederer and Stremitzer (2014).

The finding that people have a preference to conform to the expectations of someone who is socially close may have applications in the management of teams. Guilt aversion can help coordinate team members. Communicating expectations can incite team members who are socially close. At least in the realm of realistic expectations, the degree of social integration of a team can be used as a mediator to fine-tune the influence of mutual expectations. It seems like an interesting and promising avenue for future research to enrich the investigation of the impact of social closeness on guilt aversion by the impact of social status.

Finally, my results suggest that it is worth working on a truly empathic utility function that does not merely include other agents' payoffs into the utility function, but adds more elements of their utility. A theory of other-regarding reference-dependent preference with expectation-based reference points, along the lines of Köszegi and Rabin (2006), appears to be a promising starting point.

References

Abeler, J., Falk, A., Götte, L., & Huffman, D. (2011). Reference Points and Effort Provision. *The American Economic Review,* 101(2), 470-492.

Aron, A., Aron, E. N., & Norman, C. (2003). Self Expansion Model of Motivation an Cognition in Close Relationships and Beyond. In G. J. O. Fletcher & M. S. Clark (Eds.), *Blackwell Handbook of Social Psychology: Interpersonal Processes*, pp. 478–502. Blackwell Publishers Ltd.

Aron, A., Aron, E. N., Tudor, M., & Nelson, G. (1991). Close Relationships as Including Other in the Self. *Journal of Personality and Social Psychology*, *60*(2), 241–253.

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

Baumeister, R. F., Stillwell, a M., & Heatherton, T. F. (1994). Guilt: An Interpersonal Approach. *Psychological Bulletin*, *115*(2), 243–67.

Bellemare, C., Sebald, A., & Strobel, M. (2011). Measuring The Willingness To Pay To Avoid Guilt. *Journal of Applied Econometrics*, *26*, 437–453.

Bellemare, C., Sebald, A., & Suetens, S. (2013). Heterogeneous Guilt Aversion and Incentive Effects. *mimeo*

Bicchieri, C., & Chavez, A. (2010). Behaving as Expected: Public Information and Fairness Norms, *Journal of Behavioral Decision Making* 23, 161-178.

Brañas-Garza, P., Cobo-Reyes, R., Espinosa, M. P., Jiménez, N., Kovářík, J., & Ponti, G. (2010). Altruism and Social Integration. *Games and Economic Behavior*, *69*(2), 249–257.

Brewer, M. B. (2007). The Social Psychology of Intergroup Relations. *Social Psychology - Handbook of Basic Principles,* pp. 695–715.

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

Charness, G., & Dufwenberg, M. (2010). Bare Promises: An Experiment. *Economics Letters*, *107*(2), 281–283.

Charness, G., & Gneezy, U. (2008). What's in a Name? Anonymity and Social Distance in Dictator and Ultimatum Games. *Journal of Economic Behavior & Organization*, *68*(1), 29–35.

Charness, G., & Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*, *117*(3), 817–869.

Chen, Y., & Li, S. X. (2009). Group Identity and Social Preferences. *The American Economic Review*, *99*(1), 431–457.

Dawes, R. M., Van de Kragt, A. J. C., & Orbell, J. M. (1988). Not Me or Thee but We: The Importance of Group Identity in Eliciting Cooperation in Dilemma Situations: Experimental Manipulations. *Acta Psychologica*, *68*, 83–97.

Dufwenberg, M., & Gneezy, U. (2000). Measuring Beliefs in an Experimental Lost Wallet Game. *Games and Economic Behavior*, *30*(2), 163–182.

Dufwenberg, M., & Kirchsteiger, G. (2004). A Theory of Sequential Reciprocity. *Games and Economic Behavior*, *47*(2), 268–298.

Ederer, F, Stremitzer, A. (2014). Promises and Expectations. *mimeo*.

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

Engel, C. (2011). Dictator Games: a Meta Study. *Experimental Economics*, *14*(4), 583–610.

Falk, A., & Fischbacher, U. (2006). A Theory of Reciprocity. *Games and Economic Behavior*, *54*(2), 293–315.

Fehr, E., & Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, *114*(3), 817–868.

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics*, *10*(2), 171–178.

Frey, B. S., & Bohnet, I. (1999). Social Distance and Other-regarding Behavior in Dictator Games: Comment. *The American Economic Review*, *89*(1), 335–340.

Greiner, B. (2004). An Online Recruitment System for Economic Experiments. In K. Kremer & V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen – Beiträge zum Heinz-Billing-Preis 2003. GWDG-Bericht Nr. 63, Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen*, pp. 79–93.

Güth, W., Ploner, M., & Regner, T. (2009). Determinants of In-group Bias: Is Group Affiliation Mediated by Guilt-aversion? *Journal of Economic Psychology*, *30*(5), 814–827.

Hewstone, M., Stroebe, W., & Jonas, K. (2008). *Introduction to Social Psychology: A European Perspective* (4th ed.), p. 409.

Hoffman, E., McCabe, K., & Smith, V. L. (1996). Social Distance and Other-Regarding Behavior in Dictator Games. *The American Economic Review*, *86*(3), 653–660.

Köszegi, B., & Rabin, M. (2006). A Model of Reference-dependent Preferences. *The Quarterly Journal of Economics*, *121*(4), 1133–1165.

Leider, S., Möbius, M. M., Rosenblat, T., & Do, Q.-A. (2009). Directed Altruism and Enforced Reciprocity in Social Networks. *The Quarterly Journal of Economics*, (November), 1815–1851.

Leider, S., Möbius, M. M., Rosenblat, T., & Do, Q.-A. (2010). What Do We Expect From Our Friends? *Journal of the European Economic Association* 8, 120–138.

Nicole, S., Regner, T., & Harth, N. S. (2010). Other-regarding Behaviour: Testing Guilt- and Reciprocity-based Models. *mimeo.*

Ockenfels, A., & Werner, P. (in press). Beliefs and Ingroup Favoritism. *Journal of Economic Behavior & Organization*, 1–10.

Rabin, M. (1993). lncorporating Fairness into Game Theory and Economics. *American Economic Review*, *83*(5), 1281–1302.

Rankin, F. W. (2006). Requests and Social Distance in Dictator Games. *Journal of Economic Behavior & Organization*, *60*(1), 27–36.

Reuben, E., Sapienza, P., & Zingales, L. (2009). Is Mistrust Self-fulfilling? *Economics Letters*, *104*(2), 89–91.

Rockenbach, B., Böhm, R., & Weiss, A. (2013). Experimental Evidence on Identity Biases in Voting Choices, 1–27. *mimeo.*

Tajfel, H. & Turner, J. C. (1979). An Integrative Theory of Intergroup Conflict. In: W. G. Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations*. Monterey, CA: Brooks-Cole.

Vanberg, C. (2008). Why Do People Keep Their Promises? An Experimental Test of Two Explanations. *Econometrica*, *76*(6), 1467–1480.

Appendix A: Exploratory Results Regarding Unrealistic Beliefs

Predictions:

For exaggerated beliefs, this study is exploratory. In this range, conflicting forces are likely, making it difficult to derive one clear-cut hypothesis. The first force may be guilt aversion in the traditional sense. Even in the realm of exaggerated beliefs, senders may increase their amounts sent in response to increasing second-order beliefs. If beliefs are exaggerated, however, senders may also negatively condition their amounts sent on beliefs, "punishing" exaggerated beliefs. Senders may actually do so more, the more exaggerated the beliefs are. Regner and Harth (2010) found evidence that the more exaggerated beliefs are, the less trustees send in a trust game. But they explained their findings with reciprocity dominating guilt aversion in that domain. Theories of reciprocity do not – at least in their traditional form (Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006; Rabin, 1993) – apply to my design. The receiver does not act. So the sender cannot reciprocate on any kind or unkind action.

However, the "punishment" of exaggerated beliefs could also be flat. This would mean senders do discount their amount sent if beliefs are exaggerated and would send the same low amount for all excaggerated beiefs. Finally the senders would have good reason just to ignore exaggerated beliefs. All these forces may interact with a shared group identity. But again the signs of these interaction effects seem unclear. There is evidence that participants tend to be more forgiving towards people who share their group identity (Chen & Li, 2009, p. 445). But an exaggerated belief could also be interpreted as particularly presumptuous if it is stated from a group-mate meriting harsher "punishment".

Results

The regression analysis set out in the main body of the paper does not change qualitatively when including the full range of second-order beliefs, *(Random effects regression, N=120; $\beta 3=0.08$, $p=0.000$*, see below table 2, model 1). But it is obvious from plotting the average amounts sent against the whole range of second-order beliefs by treatment (Graph 1) that senders react differently to reasonable second-order beliefs than they react to exaggerated second-order beliefs. To explore the data in the realm of exagertated beliefs, I amend the original regressions by including a dummy for exaggerated beliefs I call "larger50". The dummy is equal to 1 for those amounts sent that are conditioned on the receiver's expectation that the sender sends more than half of the pie. For amounts sent conditioned on the receiver expecting not more than half of the pie, the dummy is 0. β4 denotes the coefficient of this dummy. Including this dummy confirms the observation that senders reduce the amount sent once second-order beliefs start being exaggerated (*Random effects regression, N=120; $\beta 4=-3.563$, $p=0.001$*). Adding an interaction effect of the larger50 dummy with the ingroup dummy into this regression reveals that this reduction is not stronger in statistically significant terms if senders send to outgroup members (*Random effects regression, N=120; $\beta 5=-0.346$, $p=0.875$).*

Again, the results can also be shown using the Tobit random effects model also introduced in the main body of the paper.

Table 2
All Data: second-order beliefs 0-100%

Dependent variable: **Amount sent**	Model 1: Linear regression	Model 3: Linear regression	Model 4: Linear regression	Model 2: Tobit regression	Model 5: Tobit regression	Model 6: Tobit regression
Second-order belief	0.0129 (0.0124)	0.0615** (0.0195)	0.063** (0.025)	-0.0142 (0.0221)	0.0837* (0.0347)	0.084^{+} (0.0437)
Ingroup (dummy)	-0.524 (3.371)	-0.524 (3.370)	-.445 (3.407)	-5.515 (7.204)	-5.572 (7.200)	-5.525 (7.26)
Ingroup x Second-order belief	0.0802*** (0.0175)	0.0802*** (0.0174)	0.075* (0.035)	0.157*** (0.0316)	0.158*** (0.0314)	0.155* (0.0773)
Larger50 (dummy)		-3.563** (1.105)	-3.39* (1.563)		-7.185*** (1.982)	-7.086* (2.816)
Larger50 x ingroup 1 0			-0.346 (2.211)			-0.195 (3.962)
Random effects at participant level	Yes	Yes	Yes	Yes	Yes	Yes
Constant	12.42*** (2.384)	11.61*** (2.396)	11.57*** (2.409)	-2.938 (5.166)	-4.578 (5.183)	-4.59 (5.202)
N	1320	1320	1320	1320	1320	1320
Number of groups	120	120	120	120	120	120

$^{+}$=p<0.06 *=p<0.05; **=p<0.01; ***=p<0.001
Standard errors in parenthesis

Result 5: The effect of second-order beliefs is attenuated once the realm of exaggerated beliefs is reached (Random effects regression, N=120; β4=-3.39, p<.05).
(Null) Result 6: This attenuation is not different between treatments (Random effects regression, N=120; β5=-0.346, p=.87).

To look closer at whether and, if so, how senders condition amounts sent on exaggerated beliefs, I run the regressions explained in the main body of the paper with the data on exaggerated second-order beliefs.

Table 3
Data for second-order beliefs above 50 up to 100%

Dependent variable: **Amount sent**	Model 1: Linear regression	Model 2: Tobit regression
Second-order belief	0.0005 (0.02)	0. 0005 (0.02)
Ingroup (dummy)	3.873 (0.79)	3.873 (0.8)
Ingroup x Second-order belief	0.026 (0.76)	0.026 (0.76)
Random effects at participant level	Yes	Yes
Constant	12.9*** (3.73)	12.9*** (3.75)
N	600	600
Number of groups	120	120
*=p<0.05; **=p<0.01; ***=p<0.001 Standard errors in parenthesis		

With these regressions, no significant effect can be shown. It appears that, in the realm of excessive second-order beliefs, second-order beliefs do not have any effect on the amount sent – independently of whether senders interact with ingroup or outgroup receivers.

(Null) result 6: In the realm of exaggerated beliefs, no influence of second-order beliefs on action can be shown. That is true independently of whether senders interact with ingroup or outgroup receivers (Random effects regression, N=120; β1=0.0005, p>.983; β3=0.026, p>.45).

Conclusion

In the exploratory part of the experiment, I find that exaggerated second-order beliefs generally have an attenuated influence on the amount sent. The difference of attenuation is not statistically significant between treatments. Within the realm of exggagerated beliefs, second-order beliefs seem to have no effect on the amount sent, independent of the treatment. Appendix B: Instructions in English

B.1. General Instructions (on Paper)

Welcome to our experiment!

If you read the following explanations carefully, you will be able to earn a substantial sum of money, depending on the decisions you take. It is therefore crucial that you read these explanations carefully.

During the experiment there shall be absolutely no communication between participants.

Any violation of this rule will lead to your exclusion from the experiment and from any payments. If you have any questions, please raise your hand. We will then come over to you. Please switch you mobile phone off and do not listen to music during the experiment.

In any event, you will receive a lump sum of 4 € for taking part in the experiment. During the experiment, all payoffs and earnings will be expressed in Talers. At the end of the experiment, you will be paid cash in Euro. You will receive from us the 4 € for your participation plus the sum of Talers you earned in the experiment, converted into Euros. One Taler converts to 0.11 €.

Today's experiment will consist of three parts. Before each part, the experimenter will hand you printed instructions. Now you are about to be instructed on the first phase. When the first phase is over, you will receive paper instructions on the second phase of the experiment.

B.2. Part 1 = Group Segregation Stage (Instructions on Paper)

- All participants will now be allocated to one of two groups. The groups will remain constant over the whole experiment (that is, over all three parts of today's experiment).
- The groups will be formed depending on your preferences for one of two modern painters.
- To form the two groups, everyone will be shown 5 pairs of paintings. Each pair consists of one painting by Klee and one by Kandinsky. You will not be told who painted which painting. For each pair, you will be asked to rate how much you like one painting vis-à-vis the other.
- For indicating your preference, we ask you to use a slider bar. You can move the slider bar on a continuous scale between "I strongly prefer painting L" to "I strongly prefer painting R" to indicate your preference. The middle position shows that you like both paintings equally.
- By moving the slider bar, you distribute ten points between the two paintings. The more you indicate that you like a painting, the more points are allocated to this painting and the less to the other ("I strongly prefer painting L" means 10 points for painting L; "I like both

paintings equally" means 5 points for bot paintings; "I strongly prefer painting R" means 10 points for painting R).

- The computer sums up all points you allocated to Klee and all points you allocated to Kandinsky. Then the half of the participants who allocated the most points to Klee will form the Klee group. Accordingly, the other half will form the Kandinsky group.
- As of the end of this stage, you will be informed about your group membership.
- You can read your group membership from the upper left corner of your screen at any time during today's experiment.
- The groups will remain constant over all three stages of today's experiment.

B.3. Part 2

B.3.1. Group Task (Instructions on Paper)

- You will approach this task together with the members of your group (Kandinsky or Klee). Each member will work independently, but the performance of all group members will be aggregated and constitutes a joint group performance.
- The performance of your group will be compared with the performance of the other group.
- The group with the higher performance will receive a prize of 208 Talers at the end of the experiment, which will be distributed equally among the eight group members (26 Talers per member). The group with the lower performance receives no prize. If both groups have exactly the same performance, the prize will be shared equally between the groups.
- Your task is to identify letters at certain positions in a 15-page text.
- Example: Identify the following letter: page 1, line 7, word 5, position 3. The correct solution to this example is marked in grey in the text you received – it is letter „C".
- Please indicate all letters as capitals.
- Overall you have four minutes to identify as many letters as possible.
- After four minutes, the task ends and the number of correct solutions in your group will be compared with the number of correct solutions in the other group.
- At the end of the experiment (after the third part), you will be informed which group wins the price.
- If you have read and understood the instructions, please click on "Proceed" on your computer screen.

- As soon as all participants have clicked on “Proceed”, the group task will start. Please be ready!
- Once you have completed this task, we will ask you to answer a short questionnaire, which will appear on your screen automatically.

B.3.2a. Belief Elicitation (Only on Screen, Only for Receivers)

Before we proceed with the third part of the experiment, we want you to guess the outcome of it. You will act in role B. In the experiment, you will be anonymously paired with another person who has role A. The only thing you and the person you are paired with will know about each other is to what groups you have been assigned– Kandinsky or Klee.

The person you are paired with will decide how to split 100 Talers between himself/herself and you. Every individual decision by such a person in role A will be anonymous towards both other participants and the experimenters. We want you to guess how much, on average, of the 100 Talers a person in your situation (a person in role B matched with a person of the same / different group in role A) will receive. Please enter your guess in the box below, stated in full Talers. Each participant whose guess is not more than 1 Taler off the true average amount will win 125 Talers extra, which will be paid out in the end together with whatever you earned during this experiment.

B.3.2b. Questionnaire (Only on Screen, Only for Senders).

You will now read some statements. These statements refer to the Klee [Kandinsky] group, of which you are a member. Please read the respective statement carefully and then indicate to which extent you agree with it. You can click anything between 1 (“I do not agree at all”) and 7 (“I absolutely agree”).

Example scale: I do not agree at all. 1 - 2 - 3 - 4 - 5 - 6 - 7 I absolutely agree.

1) I regard myself as a member of the Klee [Kandinsky] group.

2) I am happy about being a member of the Klee [Kandinsky] group.

3) I feel somehow connected to the members of the Klee [Kandinsky] group.

4) I identify myself as a member of the Klee [Kandinsky] group.

B.4. Part 3 = Dictator Game

- Each of you has been paired with another person in another role. You can read your role (“A” or “B”) on your screen. You will not be told who this other person is, neither during nor after the experiment.

- All you will know about this person is what group he/she belongs to (Klee/Kandinsky). You can read your own group membership on the top left of your screen at any time. You can read the group membership of the person you have been paired with from the top right of your screen at any time.

- In this part of the experiment, every person who has role A will decide how to divide 100 Talers between himself/herself and the person in role B with whom he/she has been paired. This will work as follows.

- 100 Talers each will be booked to the experimental accounts of every participant in role A.

- Every participant in role B has guessed the outcome of this experiment to be in a case like his/hers:

    - If you and the person you have been paired with are from the same group, he/she guessed the average amount a participant in role B will receive if the participant in role A he/she is paired with is for same group.

    - If you and the person you have been paired with are from different groups, he/she guessed the average amount a participant in role B will receive if the participant in role A he/she is paired with is for a different group.

- Note: This guess was made before these instructions were handed out (during the questionnaire at the end of the last part) and without the participant in role B knowing that the participant in role A he/she was paired with would be informed about the guess. Every person in role B whose guess is not more than 1 Talers off the true average will receive 125 Talers to provide an incentive to guess accurately.

- The participant in role A will now be asked what they would like to send if the receiver has guessed 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 Talers, respectively. Please fill in an answer for all these eleven cases.

- The computer will then put into effect the answer that was conditional on the belief, which is closest to the person B's actual stated guess. So, if the person in role B guessed a participant in his situation would receive 4 Talers on average, the answer for 0 Talers would be put into effect. And if the person in role B guessed a participant in his situation would receive 96 Talers on average, the answer for 100 Talers would be put into effect.

- After the experiment you will be informed about

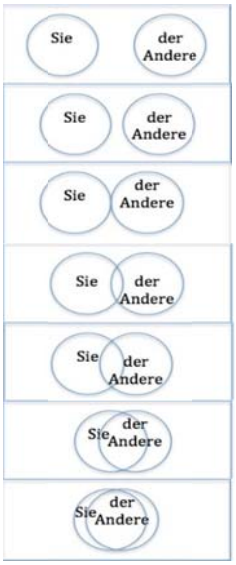
    - How much you earned in the second stage of the experiment (group competition and belief elicitation)

    - How much you earned in the third stage of the experiment (sender receiver game).

- After completion of this last part of the experiment, we would ask you please to fill in a general questionnaire while we calculate your payments. Please step forward one by one in the order of your cabin numbers as soon as the experimenter declares the experiment to be over.

B.5. Posttest

- In a posttest, participants were asked to indicate which of the pairs of circles best describes their relationship to the participant they have been paired with in the dictator game.



Appendix C: Instructions in German

Instruktionen auf Papier

Allgemeine Erklärungen für die Teilnehmer

Willkommen zu unserem Experiment!

Wenn Sie die nachfolgenden Erklärungen genau lesen, dann können Sie - je nach Ihren Entscheidungen - eine nicht unbeträchtliche Geldsumme verdienen. Es ist daher sehr wichtig, dass Sie diese Erklärungen genau durchlesen.

Während des Experiments herrscht ein absolutes Kommunikationsverbot mit den anderen Teilnehmern. Die Nichtbeachtung dieser Regel führt zum Ausschluss vom Experiment und allen Zahlungen. Wenn Sie Fragen haben, strecken Sie bitte Ihre Hand aus der Kabine. Wir kommen dann zu Ihnen.

Bitte schalten Sie Ihr Handy aus und hören Sie keine Musik.

Für Ihre Teilnahme am Experiment erhalten Sie auf jeden Fall eine Pauschale von 4 Euro.

Während des Experiments sprechen wir nicht von Euro, sondern von „Talern“. Ihr gesamtes Einkommen wird also zunächst in Talern berechnet. Die von Ihnen während des Experiments erzielte Gesamtpunktzahl wird dann am Ende in Euro umgerechnet. Dabei gilt:

1 Taler = 0.11 Euro

Am Ende bekommen Sie von uns die 4 Euro Pauschale sowie die während des Experiments verdiente Anzahl an Talern **bar** in Euro ausbezahlt.

Das heutige Experiment besteht aus drei Teilen. Vor jedem Teil werden Sie Instruktionen auf Papier erhalten. Nun wird Ihnen der erste Teil erklärt. Wenn der erste Teil vorüber ist, erhalten Sie die Instruktionen über den zweiten Teil auf Papier. Nach dem zweiten Teil erhalten Sie die Instruktionen für den dritten Teil auf Papier

Ablauf des ersten Teils

- Im ersten Teil des Experiments werden alle Teilnehmer einer von zwei Gruppen zugeteilt. Die Gruppen werden über das ganze Experiment, d.h. über alle drei Teile des Experiments, konstant bleiben.

  - Die Gruppen werden nach Ihrer Vorliebe für einen von zwei modernen Malern gebildet.

  - Um die Gruppen zu bilden, werden wir jedem von Ihnen 5 Paare von Gemälden zeigen. Jedes Paar wird aus je einem Bild von Klee und einem von Kandinsky bestehen. Sie werden nicht erfahren, welches Bild von welchem Maler stammt. Für jedes Bilderpaar werden Sie gefragt, wie sehr sie das eine Bild im Vergleich zum anderen mögen.

  - Um Ihre Vorliebe anzugeben, werden wir Sie bitten, einen Schiebe-Regler auf dem Bildschirm zu nutzen. Sie können den Schiebe-Regler auf einer kontinuierlichen Skala zwischen „Mir gefällt Bild L viel besser“ bis „Mir gefällt Bild R viel besser“ bewegen. Die Position des Schiebe-Reglers genau in der Mitte zwischen diesen beiden Enden bedeutet, dass Sie die beiden Bilder gleich stark schätzen.

  - Indem Sie den Schiebe-Regler auf der Skala bewegen, verteilen Sie zehn Punkte auf die beiden Bilder. Je mehr sie ein Bild nach Ihrer Angabe mögen, desto mehr Punkte werden diesem Bild und desto weniger Punkte werden dem anderen Bild zugeteilt

    <u>Beispiel</u>: „Mir gefällt Bild L viel besser“ bedeutet 10 Punkte für Bild L und 0 Punkte für Bild R; „Mir gefallen beide Bilder gleich“ bedeutet 5 Punkte für beide Bilder und „Mir gefällt Bild R viel besser“ bedeutet 0 Punkte für Bild L und 10 Punkte für Bild R.

  - Der Computer wird alle Punkte, die Sie den Bildern von Klee gegeben haben, aufsummieren. Ebenso wird er alle Punkte, die Sie den Bildern von Kandinsky gegeben haben, aufsummieren. Dann wird die Hälfte der Teilnehmer, die die meisten Punkte an Klee-Bilder gegeben haben, die Klee-Gruppe bilden. Entsprechend wird die andere Hälfte der Teilnehmer die Kandinsky-Gruppe bilden.

  - Am Ende dieses Teils des Experiments wird Ihnen mitgeteilt, zu welcher Gruppe Sie gehören.

  - In der oberen linken Ecke des Bildschirms werden Sie von nun an während des ganzen Experiments darüber informiert, zu welcher Gruppe Sie gehören.

  - Die Gruppen bleiben über alle drei Teile des heutigen Experiments unverändert.

Ablauf des zweiten Teils

- Die Aufgabe in diesem Teil werden Sie zusammen mit den anderen Mitgliedern Ihrer Gruppe (Klee bzw. Kandinsky) angehen. Jedes Gruppenmitglied wird unabhängig handeln, aber die Leistungen aller Mitglieder einer Gruppe werden zu einer gemeinsamen Gruppenleistung zusammengenommen.
- Die Leistung Ihrer Gruppe wird dann mit der Leistung der anderen Gruppe verglichen.
- Die Gruppe mit der besseren Leistung wird einen Preis von 208 Talern gewinnen, der gleichmäßig auf alle acht Mitglieder der Gruppe aufgeteilt wird. Die Gruppe mit der schlechteren Leistung erhält nichts. Wenn die Leistung beider Gruppen gleich ist, wird der Preis zwischen beiden Gruppen geteilt.
- Sie werden einen 15-seitigen Text erhalten und Ihre Aufgabe wird darin bestehen, Buchstaben an einer konkreten Position zu bestimmen.

  Beispiel: „Bestimmen Sie den folgenden Buchstaben: Seite 1, Zeile 7, Wort 5, Position 3.“ Die richtige Antwort zu diesem Beispiel ist in dem Text, den Sie erhalten haben, grau unterlegt – es ist der Buchstabe „C“.
- Bitte geben Sie alle Buchstaben in Großbuchstaben an.
- Insgesamt haben Sie 4 Minuten, um so viele Buchstaben zu identifizieren wie möglich.
- Nach 4 Minuten endet die Aufgabe und die Anzahl richtiger Antworten in Ihrer Gruppe wird mit der Anzahl richtiger Antworten der anderen Gruppe verglichen.
- Am Ende des Experiments (nach dem dritten Teil) werden Sie informiert, welche Gruppe die meisten richtigen Antworten gegeben und damit den Preis gewonnen hat.
- Wenn Sie diese Instruktionen gelesen und verstanden haben, klicken Sie „Weiter“ auf Ihrem Bildschirm.
- Sobald alle Teilnehmer „Weiter“ geklickt haben, wird die Aufgabe beginnen. Halten Sie sich bereit!
- Wenn diese Aufgabe beendet ist, werden wir Sie bitten, einen kurzen Fragebogen auszufüllen, der automatisch auf Ihrem Bildschirm erscheinen wird.

Ablauf des dritten Teils

- Sie wurden am Ende des letzten Teils des Experiments zufällig entweder der Rolle „A“ oder der Rolle „B“ zugeordnet.
- Welche der beiden Rollen Ihnen zugeteilt wurde, wird ihnen auf dem Bildschirm mitgeteilt, sobald der dritte Teil des Experiments beginnt.
- Jeder von Ihnen wurde mit jeweils einer Person in einer anderen Rolle gepaart.
- Sie werden weder während noch nach dem Experiment erfahren, wer die Person ist, mit der Sie gepaart wurden.
- Alles, was Sie über diese Person wissen werden, ist, welcher Gruppe (Klee oder Kandinsky) sie angehört.
- Sie können Ihre eigene Gruppenzugehörigkeit jederzeit von der linken oberen Ecke Ihres Bildschirms ablesen. Sie können die Gruppenzugehörigkeit der Person, mit der Sie gepaart wurden, jederzeit von der rechten oberen Ecke Ihres Bildschirms ablesen.
- In diesem Teil des Experiments wird jede Person in der Rolle „A“ entscheiden, wie sie 100 Taler zwischen sich selbst und der Person, mit der sie gepaart ist, aufteilen wird. Dies wird wie folgt funktionieren.
- 100 Taler werden jeweils auf das Konto jeder Person in Rolle „A“ gebucht.
- Jeder Teilnehmer in der Rolle „B“ hat geschätzt, wie diese Aufteilung in einem Fall wie dem Ihren ausgehen wird:
    - Falls Sie und die mit Ihnen gepaarte Person aus der selben Gruppe sind, hat die Person in Rolle „B“, folgende Frage beantwortet: Wie viel wird eine Person in Rolle „B“ im Durchschnitt erhalten, wenn sie mit einer Person in Rolle „A“ gepaart wurde, die aus der selben Gruppe (Klee/Kandinsky) stammt wie sie selbst.
    - Falls Sie und die mit Ihnen gepaarte Person aus unterschiedlichen Gruppen stammen, hat die Person in Rolle „B“, folgende Frage beantwortet: Wie viel wird eine Person in Rolle „B“ im Durchschnitt erhalten, wenn sie mit einer Person in Rolle „A“ gepaart wurde, die aus einer anderen Gruppe (Klee/Kandinsky) stammt als sie selbst.
- Beachten Sie: Diese Schätzung hat die Person in Rolle „B“ gemacht, bevor diese Instruktionen ausgeteilt wurden (während der Fragebogenphase im letzten Teil des Experiments). Die Teilnehmer in Rolle „B“ wussten nicht, dass der Teilnehmer in Rolle

„A“ über ihre Schätzung informiert wird. Jede Person in Rolle „B“, deren Schätzung nicht mehr als einen Taler vom wirklichen Durchschnitt entfernt liegt, wird 125 Taler erhalten. Das sollte einen Anreiz bieten, eine zutreffende Schätzung abzugeben.

- Die Teilnehmer in Rolle „A“ werden nun gefragt, welchen Betrag sie dem jeweiligen Teilnehmer in Rolle „B“ senden wollen, falls der Teilnehmer in Rolle „B“ 0, 10, 20, 30, 40, 50, 60, 70, 80, 90 bzw. 100 Taler geschätzt hat. Bitte antworten Sie für **jeden** dieser Fälle.

- Der Computer wird dann die Antwort für den Fall umsetzen, der am nächsten an der wirklichen Schätzung der Person in Rolle „B“ liegt.

  Beispiel: Falls die Person in Rolle „B“ geschätzt hätte, dass eine Person in ihrer Situation im Durchschnitt vier Taler empfangen würde, würde der Computer die Antwort der Person in Rolle „A“ für den Fall „0 Taler“ umsetzen. Falls die Person in Rolle „B“ geschätzt hätte, dass eine Person in ihrer Situation im Durchschnitt 96 Taler erhalten würde, würde der Computer die Antwort der Person in Rolle „A“ für den Fall „100 Taler“ umsetzen

- Hiernach stellen wir Ihnen noch eine Frage.

- Danach werden Sie darüber informiert,

  - wie viel Sie im zweiten Teil des Experiments verdient haben (Gruppenaufgabe und Schätzung)

  - wie viel Sie im dritten Teil des Experiments verdient haben (Sender-Empfänger-Aufgabe)

- wie viel Sie insgesamt verdient haben.

- Bitte beantworten Sie nach dem letzten Teil des Experiments noch einen allgemeinen Fragebogen, während wir Ihre Auszahlung berechnen. Bitte kommen Sie dann einzeln in der Reihenfolge Ihrer Kabinennummern zum Auszahlungstisch, wenn der Leiter des Experiments Sie dazu auffordert.

Beliefabfrage (nur auf dem Bildschirm, nur für Empfänger)

Bevor wir zur dritten Phase des Experiments kommen, möchten wir Sie bitten, das Ergebnis der dritten Phase vorherzusagen.
Dazu erklären wir Ihnen schon hier, was in der dritten Phase des Experiments geschehen wird.
Ihnen ist für die dritte Phase die Rolle des B-Spielers zugelost worden. Ferner wurden Sie für die dritte Phase schon zufällig einem anderen Teilnehmer in der Rolle A zugeordnet. Die einzige Information, die Sie über den anderen Teilnehmer haben werden, ist, ob dieser zur

Klee- oder zur Kandinsky-Gruppe gehört. Ebenso wird er über Sie nur wissen, ob Sie zur Klee- oder Kandinsky-Gruppe gehören.
Sie gehören zur Kandinsky-Gruppe [Klee-Gruppe].
Der Ihnen zugeteilte Teilnehmer in Rolle A gehört zur Kandinsky-Gruppe [Klee-Gruppe].

Der Teilnehmer in Rolle A, dem Sie zugeordnet sind, wird darüber entscheiden, wie 100 Taler zwischen Ihnen beiden aufgeteilt werden.
Wir bitten Sie, zu schätzen, wie viel von den 100 Talern ein Teilnehmer in Rolle B im Durchschnitt von einem Teilnehmer in Rolle A erhalten wird, wenn er sich in einer Paarung wie der Ihren befindet (beide Teilnehmer gehören der selben Gruppe an [die beiden Teilnehmer gehören unterschiedlichen Gruppen an]).
Jeder Teilnehmer, dessen Schätzung ausreichend präzise ist, erhält 125 Taler, die am Ende des Experiments zusammen mit dem Betrag ausgezahlt werden, den er während des Experiments verdient hat. Ausreichend präzise ist Ihre Schätzung, wenn sie nicht mehr als 1 Taler vom „wahren Durchschnittswert“ entfernt liegt. Der „wahre Durchschnittswert“ ist der durchschnittliche Betrag, den Teilnehmer in Rolle B in einer Paarung wie der Ihren (beide Teilnehmer gehören derselben Gruppe an [die beiden Teilnehmer gehören unterschiedlichen Gruppen an]) in dieser Session von Teilnehmern in Rolle A erhalten.
Bitte geben Sie Ihre Schätzung in das Feld unter diesem Text ein. Sie können jeden Betrag zwischen 0 und 100 Talern angeben, der sich in ganzen Talern ausdrücken lässt.
Ein Teilnehmer in Rolle B der - wie in Ihrem Fall – zur selben Gruppe gehört wie [zu einer anderen Gruppe gehört als] der Teilnehmer in Rolle A, der ihm zugeordnet ist, erhält im Durchschnitt ... Taler.

Fragebogen auf dem Bildschirm für die Sender).

Sie werden nun einige Aussagen lesen. Die Aussagen beziehen sich auf die Klee [Kandinsky] -Gruppe, der Sie angehören. Lesen Sie sich die jeweilige Aussage genau durch und klicken Sie an, wie sehr Sie ihr zustimmen oder nicht zustimmen. Dabei können Sie zwischen 1 "stimme überhaupt nicht zu" und 7 "stimme absolut zu" wählen.
Beispielskala: stimme überhaupt nicht zu 1 - 2 - 3 - 4 - 5 - 6 - 7 stimme voll zu
1) Ich sehe mich selbst als Mitglied der Klee [Kandinsky]-Gruppe
2) Ich bin froh darüber, zur Klee [Kandinsky]-Gruppe zu gehören.
3) Ich fühle mich irgendwie den anderen Mitgliedern der Klee [Kandinsky]-Gruppe verbunden.
4) Ich identifiziere mich als Mitglied der Klee [Kandinsky]-Gruppe.

Posttest

Im Posttest wurden die Teilnehmer gebeten, das Kreispaar anzuklicken, das am ehesten ihre Beziehung zu dem mit ihnen für das Diktatorspiel gepaarten Teilnehmer beschrieb.