Games and Economic Behavior 68 (2010) 95–107

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Testing guilt aversion

Tore Ellingsen[a,*], Magnus Johannesson[a], Sigve Tjøtta[b], Gaute Torsvik[b]

[a] Department of Economics, Stockholm School of Economics, Box 6501, SE-113 83 Stockholm, Sweden
[b] Department of Economics, University of Bergen, Hermann Fossgate 6, 5007 Bergen, Norway

ARTICLE INFO

Article history:
Received 6 May 2008
Available online 7 May 2009

JEL classification:
C91
D64

Keywords:
Guilt aversion
Beliefs
Generosity
Experiments

ABSTRACT

Guilt averse individuals experience a utility loss if they believe they let someone down. For example, generosity depends on what the donor believes that the recipient expects to receive. We measure guilt aversion in three separate experiments: a dictator game experiment, a complete information trust game experiment, and a hidden action trust game experiment. In the experiments we inform donors about the beliefs of the matched recipients, while eliciting these beliefs so as to maximize recipient honesty. The correlation between generous behavior and elicited beliefs is close to zero in all three experiments, suggesting that guilt aversion is smaller than previously thought.

1. Introduction

How should economists model generous and trustworthy behavior? A growing body of evidence suggests that generosity and trustworthiness are not simply features of people's preferences over consumption allocations, or even over transfers.[1] Social context and symbolic meaning also matter. Specifically, people are concerned about others' *beliefs*. "What is expected of me?" "What will others think of me?"

The *guilt aversion model* addresses the first of these concerns, postulating that people may feel guilty if their behavior falls short of others' expectations; see Charness and Dufwenberg (2006) for an intuitive account and Battigalli and Dufwenberg (2007, 2009) for formalities.[2]

Guilt aversion is quite hard to identify empirically. Guilt depends on one's beliefs about others' anticipations, and both anticipations and beliefs about anticipations are difficult for the researcher to manipulate. Therefore, the existence of guilt aversion has usually been inferred by asking subjects what they believe that the opponents expect; see Bacharach et al. (2007), Charness and Dufwenberg (2006), Dufwenberg and Gneezy (2000), and Guerra and Zizzo (2004).[3] The studies reveal significant correlations between second-order beliefs and actions. The evidence is clearly consistent with the guilt aversion hypothesis. However, the test may be quite weak. As Charness and Dufwenberg (2006, p. 1594) note in the discussion of their trust game evidence, an alternative explanation is that trustees' behavior anchor on their own guesses concerning

* Corresponding author.
E-mail addresses: tore.ellingsen@hhs.se (T. Ellingsen), magnus.johannesson@hhs.se (M. Johannesson), sigve.tjotta@econ.uib.no (S. Tjøtta), gaute.torsvik@econ.uib.no (G. Torsvik).

1 Pure altruism, as analyzed by Becker (1974), entails preferences over allocations only. Impure altruism, as analyzed by Andreoni (1989, 1990), also admits preferences over transfers.

2 Battigalli and Dufwenberg (2007, 2009) build on the "psychological game theory" framework pioneered by Geanakoplos et al. (1989) (GPS) and used by Rabin (1993) to study fairness issues.

3 For a similar test in a public goods game, see Dufwenberg et al. (2006).

0899-8256/$ – see front matter
doi:10.1016/j.geb.2009.04.021

trustor expectations. A closely related confound is that trustees may believe that other trustees would choose like them, and think that trustors' beliefs lean in this direction too. Trustees preferring to make larger back-transfers to the trustor are then exactly those trustees who believe that trustors expect large back-transfers. The belief that others' perceptions and thoughts are similar to one's own is known as the (false) consensus effect (Ross et al., 1977).[4]

In the present paper, we attempt to test the guilt aversion hypothesis in a way that reduces the scope for consensus effects. The tests work by providing information to each player about the paired player's first-order beliefs.[5] Our simplest experiment elicits recipient beliefs in a Dictator game and communicates these beliefs to the dictator before the dictator makes the allocation decision.[6] The guilt aversion hypothesis postulates a positive relationship between the elicited belief and the amount of money allocated to the recipient, and our test does not admit a reverse causality interpretation.

If recipients were to realize that their beliefs will be communicated to the dictator, reported beliefs may be untruthful. Indeed, if recipients believe that dictators are guilt averse, reports should be distorted upwards, potentially heavily so. In order to eliminate or at least mitigate such strategic reporting, we do not tell recipients that the dictators will have access to their beliefs. We merely ask them to guess the outcome (and even provide a material incentive to guess accurately).[7]

Our dictator game evidence reveals virtually zero correlation between the recipient's guess and the dictator's allocation decision. Since the Dictator game depicts just one (quite extreme) social situation, it would be premature to reject the guilt aversion hypothesis on the basis of this experiment alone. Therefore we also conduct two different trust game experiments, the latter of which is virtually identical with the hidden action trust game considered by Charness and Dufwenberg (2006). In these games, we again elicit trustors' expectations and investigate to what extent they affect trustees' behavior. As in the Dictator game, the correlation is close to zero. Only when we restrict attention to a subset of subjects in the first of the two trust games do we find patterns that may indicate guilt aversion. We conclude that guilt aversion appears to play a relatively minor role in these experiments, and that previous findings to the contrary may be driven by consensus effects.[8]

Before discussing the evidence in more detail, let us address the possible objection that our design violates the norm of experimenter honesty. The concern is that we do not tell subjects that their guesses will be communicated to their counterpart. We think that the design complies with the honesty norm for the following three reasons. First, and most importantly, we never lie to the subjects. Second, the honesty norm does not imply telling subjects everything. Quite the opposite: a major concern in most experimental research is to minimize "experimenter demand effects." In order to prevent subjects from doing what they believe the experimenter desires, subjects are usually not informed about the purpose of the study, even though (or precisely because) such information would affect their behavior. Not telling everything is thus acceptable in and of itself. Having said this, we acknowledge that there is a difference between subjects being ignorant about the experimenter's purpose and subjects holding erroneous beliefs about what is happening in the experiment. Third, our design in no way compromises subject integrity. To the contrary, not telling subjects that their guesses will be communicated removed an incentive for them to behave strategically by reporting a false guess. In post-experimental conversations, no subject reported feelings of being cheated or had any other misgivings about the experiment. This being said, we cannot deny a referee's concern that those subjects who learned the guesses of their opponents might come to think that "perhaps not everything is what it seems." If these subjects take part in future experiments, they may be more alert to hidden possibilities. Such heightened awareness could create a potential negative externality on other experimenters drawing subjects from this pool.[9]

Even if the design complies with experimental norms, we clearly cannot rule out that our secret sharing of information may have had an independent effect on the informed subjects' behavior. However, as long as there is no interaction between this effect and the effect of the guess, our test of the guilt aversion hypothesis remains valid.

[4] Dawes (1989, 1990) observes that knowledge of oneself does potentially provide a signal about others, so not all consensus effects are false. For evidence on true and false consensus effects, see Engelmann and Strobel (2000, 2007). As a referee has noted, our argument actually utilizes the consensus effect twice: Trustees believe that other trustees *behave* likewise and that trustors *think* likewise.

[5] Bacharach et al. (2007) and Guerra and Zizzo (2004) have previously elicited predictions from trustors. However, they informed trustees only about the mean prediction made by trustors with whom they were not matched. They do not report correlations between trustee behavior and this average.

[6] Dufwenberg and Gneezy (2000) also used a Dictator game to investigate guilt aversion. While they elicited first-order beliefs from recipients, these were not communicated to dictators in any form. As in Charness and Dufwenberg (2006) only second-order beliefs from dictators were used to measure guilt aversion.

[7] A referee has suggested that the recipients may have anticipated that guesses would be handed on to dictators, or at least that dictators suspected such anticipation. We have no indications of recipient anticipation, but cannot rule out dictator suspicion. The literature on information projection (e.g., Camerer et al., 1989; Madarász, 2008) shows that people often under-appreciate that others lack the information that they themselves have access to. If a substantial fraction of dictators believes that messages are strategic, there is less scope for finding evidence of guilt aversion.

[8] Independently of us, Vanberg (2008) has reached a related conclusion. He investigates the hypothesis that people keep promises because (i) promises affect expectations and (ii) people feel guilty when not fulfilling these expectations. In his experiment, Vanberg finds that people frequently keep their own promises, but are not particularly prone to keep the promises of others, even if only the promisors themselves know who made the promise. Since all promises are associated with an increase in second-order beliefs, but promise-keeping is not, Vanberg's findings also question the guilt aversion hypothesis, at least in its current form.

[9] In the current case, the externalities are small. First, these subjects do not participate in many experiments, and will typically not be exposed to other experimenters than us. Second, direct contamination across experiments is unlikely because opportunities for information leakage only rarely arise in economic experiments. Third, if subjects have unjustified concerns about information leakage in future experiments, the experimenter might simply clarify that information will not leak.

2. Experiment I

The first experiment is a double-blind dictator game, in which neither other subjects nor the experimenters can observe the decision of a particular subject (Hoffman et al., 1994, 1996).[10]

2.1. Design

One subject (the dictator) decides how to allocate SEK 120 between herself and another subject (the recipient) in another room (SEK = Swedish Kronor; $1 ≈ SEK 8 at the time of the experiment).

To identify the effect of an exogenous variation in second-order beliefs, we elicit recipients' beliefs about donations prior to the dictator game. Every recipient guesses how much dictators will give on average, and the best guess is rewarded a SEK 100 payment after the experiment.[11,12] Prior to making the allocation decision, the dictator is shown the guess of the matched recipient. Let us call this guess *the induced second-order belief*.

To counter the risk that recipients' guesses are contaminated by a desire to influence the dictator, recipients are not told that their guess will be shown to the dictator.[13] If second-order beliefs are important for donations as predicted by guilt aversion, there should be a positive correlation between the dictator allocation and the recipient's guess.[14]

The subjects were recently enrolled undergraduate students at the Stockholm School of Economics in Sweden. We conducted three sessions of the experiment and a total of 171 subjects participated. Three of these were used as monitors (see Appendix A). The remaining 168 subjects yielded 84 pairs of observations. For a fuller description of the experimental procedures and a complete set of instructions, see Appendix A.

2.2. Results

The average donation is 24% of the endowment. Almost two thirds (65%) of the subjects donate something, and nearly 20% of the subjects split the endowment 50/50. Perhaps surprisingly, four subjects (5% of the subjects) donate the entire amount.

The most common beliefs are 17% (SEK 20) and 33% (SEK 40), followed by 50% (SEK 60). On average the recipients expect to get 32% of the endowment, which is higher than the mean donation of 24% ($p = 0.015$ according to an independent samples t-test; according to a Mann–Whitney test, the distribution of guesses is likewise above the distribution of donations, with $p = 0.002$).

In Fig. 1 we plot the relationship between donations and beliefs. It is difficult to spot any clear-cut pattern between recipient beliefs and actual donations. The lack of connection is confirmed by the correlation coefficients. The parametric Pearson correlation coefficient is -0.075 and far from significant ($p = 0.497$); the non-parametric Spearman correlation coefficient is -0.044 ($p = 0.689$). As a sensitivity analysis we removed all 29 subjects who donated nothing – and for whom guilt aversion may be a secondary concern. The Pearson correlation coefficient is then -0.026 ($p = 0.852$) and the Spearman is 0.091 ($p = 0.509$). We also removed the six observations with beliefs exceeding 50%. The correlation coefficients then approach zero (-0.008 (Pearson) and -0.007 (Spearman)). We finally tried removing the four subjects who donated everything, but also in this case the correlation coefficients are close to zero (0.012 (Pearson) and 0.012 (Spearman)).[15] We therefore cannot reject the null hypothesis of a zero correlation. Despite the relatively large number of subjects and the associated statistical power, we fail to find support for guilt aversion.

[10] A difference compared to previous double-blind dictator game experiments is that no dictator received an envelope without any money (this procedure has previously been used as an extra guarantee of anonymity, as even if no dictator donates any money the experimenter will be unable to infer the decision of a specific subject).

[11] In principle, the optimal guess may differ slightly from one's expectation of the average, since the probability of winning depends on the distribution of other guesses. However, even if subjects would have identical expectations the theoretical effect is minor, and we see no evidence suggesting that it is important in practice; for example, all subjects except one make integer guesses, which they should not if they worried about ties. At any rate our test primarily requires that the guess is positively correlated with the true belief. (Note also that our other two experiments avoid the whole issue by rewarding guesses based on absolute rather than relative performance.)

[12] Paying subjects for correct guesses introduces a possible bias if subjects use the stated guess to hedge their income from the experiment; see Blanco et al. (2008). For example, a recipient could falsely state a belief that donations would be small in order to get some money in case the donation is actually small. Since recipients guess the average rather than the own received amount, the problem is mitigated. Moreover, stakes are modest, so the demand for hedging is probably small to begin with. Finally, for us hedging is not a problem as such; it is only problematic if donors believe that recipients hedge instead of stating their true belief.

[13] The dictator is also informed that the recipient made the guess without knowing that it would be shown to the dictator.

[14] Anchoring provides an alternative hypothesis for why we would observe a positive correlation in the experiment. Research in psychology suggests that subjects may anchor on numbers or monetary amounts even if they provide no valuable information (Tversky and Kahneman, 1974; Ariely et al., 2003). To assess the size of any anchoring effects we therefore also conducted a control treatment (also with separate subjects), in which the recipients participated in a guessing game prior to the dictator game. They were asked to guess a randomly drawn amount between SEK 0 and SEK 120, and the closest guess was rewarded a SEK 100 payment. These guesses were shown to the dictators before they made their allocation decisions. If there is an anchoring effect, the dictator donations will be correlated with these guesses. We found no evidence of anchoring.

[15] If we remove both the four subjects that donated everything and the six subjects that received beliefs over 50%, the correlation coefficients are positive (0.099 (Pearson) and 0.051 (Spearman)), but still far from significant ($p = 0.400$ and $p = 0.668$, respectively).

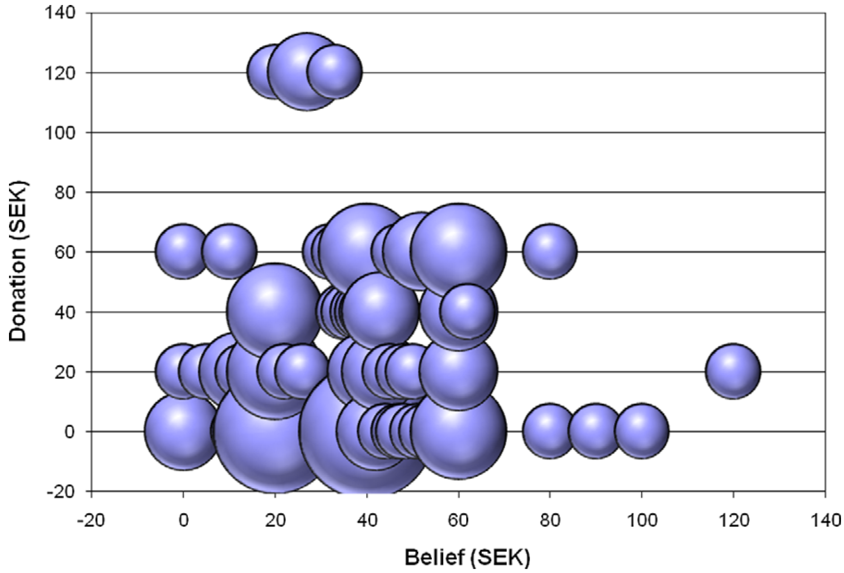

Fig. 1. A bubble plot of donations and beliefs in Experiment I (the dictator game experiment).

3. Experiment II

The second experiment investigates the effect of induced second-order beliefs in a double-blind trust game.[16] As indicated in the Introduction, a substantial fraction of the prior evidence on guilt aversion comes from trust game experiments; see Bacharach et al. (2007), Dufwenberg and Gneezy (2000), and Guerra and Zizzo (2004).

3.1. Design

At the first stage of the game one subject, the trustor, decides whether to send NOK 50 to a second player, the trustee (NOK = Norwegian Kroner; $1 ≈ NOK 6 at the time of the experiment). If the sum is sent, it is multiplied by 5. In the second stage, the trustee decides how much of the total sum of 250 to send back to the trustor (the back-transfer), but is restricted to choose an amount from the set $\{0, 50, 100, 150, 200, 250\}$. Note that the second stage is essentially a dictator game.[17]

To identify the effect of an exogenous variation in second-order beliefs we elicit trustors' beliefs about back-transfers after the trustors have made their investment decision. Every trustor guesses the most common amount a trustee will back-transfer conditional on investment (i.e., they guess the mode of the distribution). A correct guess is rewarded with NOK 50. The trustee is shown the guess of the trustor prior to making the allocation decision. As in Experiment I, we did not inform the trustors that the trustees would observe their guesses. If (induced) second-order beliefs are important for back-transfers as predicted by guilt aversion, there should be a positive correlation between the trustor belief and the back-transfer of the trustee.[18]

The subjects were students at the University of Bergen in Norway. We conducted four sessions. A total of 196 subjects participated in the experiment, yielding 98 subject pairs. As we are interested in the behavior of the trustees, we lose observations for the ten pairs where no investments were made, leaving us with 88 observations in the analysis. For more detailed experimental procedures and complete instructions, see Appendix B.

3.2. Results

The average back-transfer is 43% of the endowment. The most common back-transfers are 40% (NOK 100) or 60% (NOK 150) of the endowment. The most common beliefs are that the trustee will transfer 20%, 40%, or 60%. The aver-

[16] Like Camerer (2003) we define trust games quite broadly, including both the classical investment game of Berg et al. (1995) and the various simplifications of this game in which each player has fewer (but at least two) actions to choose between. Other notable trust game experiments include (but are not limited to) Cox (2004), Cox and Deck (2005), Dufwenberg and Gneezy (2000), and McCabe et al. (2003).

[17] The experiment is similar to the f4 experiment in Dufwenberg and Gneezy (2000). However, note that the trustee makes the back-transfer decision after learning the investment decision. Dufwenberg and Gneezy instead use the strategy method.

[18] As for Experiment I, we include a control treatment to control for any anchoring effects. In the control treatment trustors participate in a guessing game prior to the trust game where they are asked to guess on one of the following randomly drawn amounts: NOK 0, NOK 50, NOK 100, NOK 150, NOK 200, NOK 250. Correct guesses are rewarded a NOK 50 payment. These guesses are shown to the trustees before they make their back-transfer decision. If there is an anchoring effect the back-transfers will be correlated with these guesses. We found no evidence of an anchoring effect.

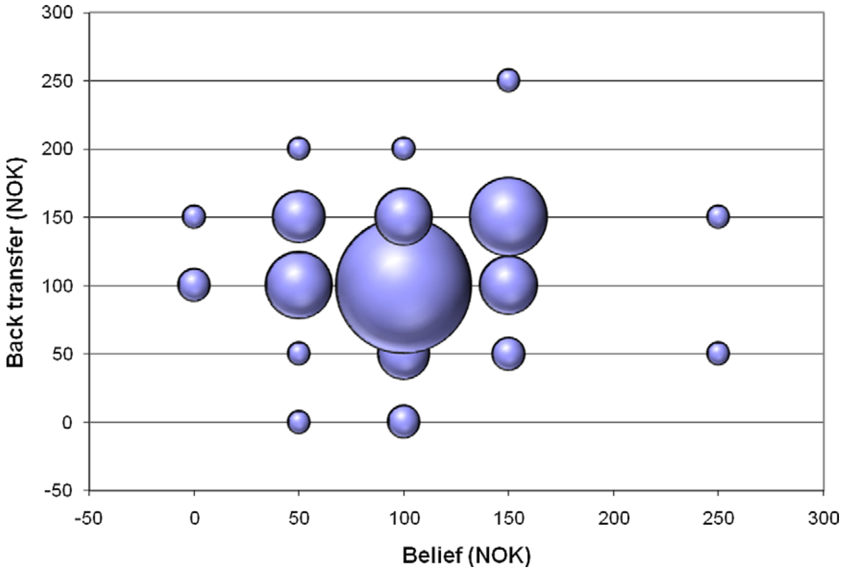

Fig. 2. A bubble plot of back-transfers and beliefs in Experiment II (the trust game experiment).

age belief is 41%, which is not significantly different from the mean back-transfer ($p = 0.287$ according to an independent samples t-test; the distributions are likewise insignificantly different, with $p = 0.194$ according to a Mann–Whitney test).

In Fig. 2 we plot the relationship between back-transfers and beliefs. There is no clear-cut pattern between beliefs and back-transfers in Fig. 2. The parametric Pearson correlation coefficient is 0.085 ($p = 0.434$) and the non-parametric Spearman correlation coefficient is 0.124 ($p = 0.249$). Removing the subject that transferred everything has little effect on the results (Pearson correlation = 0.044 ($p = 0.683$) and Spearman correlation = 0.096 ($p = 0.378$)).[19] Removing the two subjects with beliefs of a 100% back-transfer also yields similar results (Pearson correlation = 0.122 ($p = 0.264$) and Spearman correlation = 0.140 ($p = 0.197$)).[20] It thus appears difficult to reject the null hypothesis of a zero correlation.

However, as noted by a referee, there may be a case for focusing particular attention on beliefs 100 and 150. These are the two beliefs that perhaps make most sense for an investing trustor: A back-transfer of 100 is the smallest amount that yields a positive private return, and 150 is the only other amount that comes close to the equal split of the net return, $250/2 = 125$. If we restrict attention to these two beliefs only, there is a positive correlation between beliefs and back-transfers (Pearson correlation = 0.354 ($p = 0.003$) and Spearman correlation = 0.380 ($p = 0.001$)). Even if it involves a bit of stretching, this evidence may be suggestive of guilt aversion.

4. Experiment III

We finally investigate the presence of guilt aversion in a trust game with hidden actions, as developed by Charness and Dufwenberg (2006). For ease of comparison, we describe the game using the original dollar payoffs.

4.1. Design

There are two players (A and B). In the first stage, player A decides between In and Out. If Out is chosen the game ends and both players receive \$5. If player A chooses In, player B decides between Roll and Don't Roll. If Don't Roll is chosen, player B receives \$14 and player A receives nothing. If Roll is chosen player B always receives \$10, but rolls a six-sided die to determine the payoff for player A (with probability $1/6$ A receives nothing and with probability $5/6$ A receives \$12). Player A is not informed about B's decision, but A can infer that B chose Roll if the payoff is \$12. We use the same relative payoffs, but multiply by 20 to get the payoffs in NOK.

To measure guilt aversion, Charness and Dufwenberg let A players guess what fraction of B players will chose Roll. To measure second-order beliefs they then let B players guess the beliefs of A, conditional on A selecting to move In. To get an exogenous variation in beliefs we modify this design so that B players instead are informed about the guess (belief) of their paired A player before deciding whether to Roll or not. As Charness and Dufwenberg, we use the strategy method, so that all B players choose between Roll and Don't Roll before they are informed about whether their counterpart chose In or Out. In our experiment player A first choose In or Out. Player A then guesses what fraction of B players will chose Roll

[19] If we also remove the two subjects that transferred 80%, the Pearson correlation is 0.086 ($p = 0.432$) and the Spearman correlation is 0.138 ($p = 0.209$).

[20] If we remove both the subject transferring 100% and the two subjects with beliefs of 100% back-transfers, the Pearson correlation coefficient is 0.072 ($p = 0.512$) and the Spearman correlation coefficient is 0.109 ($p = 0.319$).

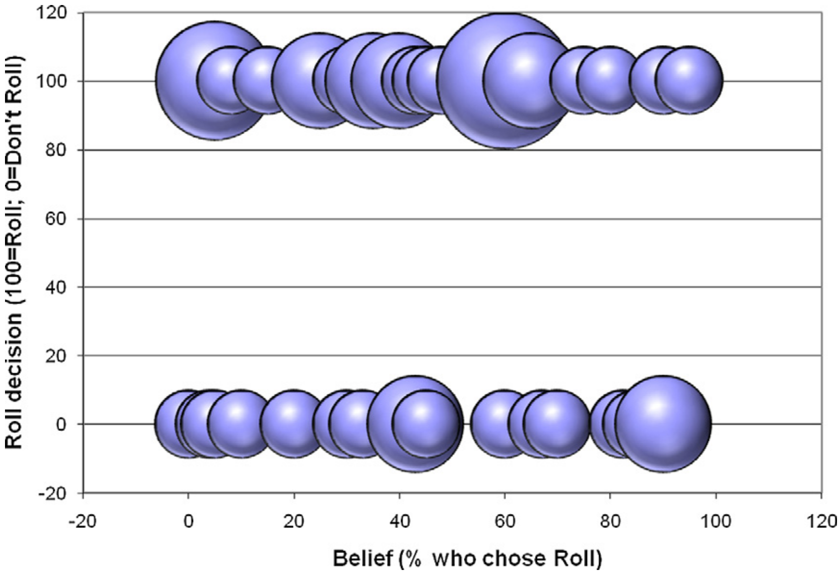

Fig. 3. A bubble plot of player B's Roll decisions and beliefs in Experiment III (the trust game with hidden action experiment).

(without being informed that this guess will be given to player B). If the guess is within 5 percentage units of the actual fraction, player A earns NOK 50. Player B is informed about the guess of player A, and thereafter decides whether to choose Roll or Don't Roll.[21] Finally, player B is informed about whether player A chose In or Out. According to the guilt aversion hypothesis, there should be a positive correlation between choosing Roll and the induced second-order belief (the guess of player A).[22]

The subjects were students at the University of Bergen in Norway. A total of 88 subjects participated in the experiment, yielding 44 subject pairs. The sample size is thus similar to the sample size in Charness and Dufwenberg, which had 45 observations in the corresponding treatment. For more detailed experimental procedures and complete instructions, see Appendix C.

4.2. Results

At the first stage, 59% of A's choose In. At the second stage, 61% of B's choose Roll. The combination of In and Roll occur in 34% of the subject pairs. These fractions are quite similar to those of Charness and Dufwenberg.

The average belief is 43%, which is lower than the actual rate of Roll.[23] Charness and Dufwenberg did not report correlation coefficients between beliefs and the decision to Roll, but compared the difference in beliefs between those who chose to Roll and those who chose Don't Roll. For the treatment that is closest to ours (no communication and (5.5) as Out payoffs), they reported mean beliefs of 54.2% for Roll and 39.6% for Don't Roll. For the same comparison, we obtain mean beliefs of 41.6% for Roll and 45.7% for Don't Roll. The latter two numbers are not significantly different ($p = 0.657$ according to an independent samples t-test; the distributions are also similar according to a Mann–Whitney test, with $p = 0.664$).

In Fig. 3 we plot the relationship between beliefs and the decision to Roll. There is no immediately discernable pattern. The Pearson correlation coefficient is -0.071 ($p = 0.646$) and the Spearman correlation coefficient is -0.066 ($p = 0.669$). Removing the highest and lowest beliefs (95% and 0%) entails similar results; the Pearson correlation coefficient is -0.162 ($p = 0.306$) and the Spearman correlation coefficient is -0.160 ($p = 0.311$). Since coefficients have negative signs, our results if anything run counter to the guilt aversion hypothesis.

The only data point that is somewhat suggestive of guilt aversion is the trustee response to trustor beliefs of a 60% chance of Roll. A remarkably large fraction (more than 11%) of trustors holds this belief. As is evident from Fig. 3, trustees tend to Roll when faced with this belief. In our discussion of Experiment II, we mentioned that there is some support for guilt aversion for a certain set of "reasonable" beliefs. Is the same true here? Are trustees guilt averse when facing those beliefs which are most "reasonable" for trustors who chose In? We think not. If we restrict attention to beliefs in the range 60–100%, the Roll rate is again negatively related to the belief; the Pearson correlation coefficient is -0.268 ($p = 0.299$) and the Spearman correlation coefficient is -0.297 ($p = 0.246$).

[21] Player B is also informed that player A did not know that the guess would be shown to player B.

[22] Observe that a trustee receives the trustor's guess before knowing whether the trustor has chosen In. This is necessary in order to utilize the strategy method for the trustees, as do Charness and Dufwenberg.

[23] This difference in means is significant at the 5% level with an independent samples t-test ($p = 0.038$) and the distributions are different at the 10% level according to a Mann–Whitney test ($p = 0.052$).

One driver of our result is the behavior of trustees who believe trustors to be strongly pessimistic. For example, we find that a majority of trustees faced with a stated belief of less than 10% chooses Roll. While it is possible that these trustees choose Roll because they believe it is unlikely that their matched trustor will have chosen In (in which case it is cheap to do the right thing), a similar argument also applies in Charness and Dufwenberg's setting.[24] To us, the most plausible interpretation of the difference is therefore that in our experiment it is random which trustee is faced with such pessimistic beliefs, whereas in Charness and Dufwenberg's experiment the trustees who believe trustors to be pessimistic are precisely those trustees who are least inclined to Roll.

As noted by a referee, it is not entirely clear that the differences between our experiment and previous experiments are due to consensus effects. A competing hypothesis is that Scandinavian subjects are less guilt averse than American (and British) subjects. To check whether this is the case, we replicate Charness and Dufwenberg's design, recruiting 94 new subjects (47 pairs) from our Norwegian subject pool. We find that trustees who choose Don't Roll have average second-order beliefs of 51.4%, whereas trustees who choose Roll have average second-order beliefs of 65.5%. (Note that in this replication treatment we ask what trustees believe that *trustors who choose In believe*, since this is what Charness and Dufwenberg do. This probably accounts for part of the difference between the overall actual trustor belief of 43% in our original treatment.) Recall that in Charness and Dufwenberg's study, the corresponding numbers are: 45 pairs, 39.6% and 54.2%. Thus, average beliefs in our subject pool are higher than in theirs, but the absolute difference between the two groups of trustees is quite similar, at about fourteen percentage points.[25] We conclude that the discrepancies in guilt aversion measures between our study and Charness and Dufwenberg (2006) are caused primarily by the differences in experimental designs rather than differences in subject pools.

5. Final remarks

When decision makers have information about opponents' beliefs, our measure of guilt aversion is insignificantly different from zero in each of our three experiments. The experiments are relatively large, involving a total of 455 subjects in two countries – not counting the several hundred participants taking part in the various control treatments (cf. footnotes 14 and 18, and the previous paragraph). Our findings suggest that consensus effects are responsible for a substantial fraction of the correlation between second-order beliefs and behavior in other studies, and that guilt aversion is accordingly smaller than previously thought.

Acknowledgments

Thanks to Gary Charness, Martin Dufwenberg, Robert Östling, and three referees for helpful comments on earlier drafts and to Anna Breman, Eirik Fjellså Hærem, Stein Svalestad and Niklas Zethraeus for research assistance. Financial support from The Jan Wallander and Tom Hedelius Foundation, the Swedish Research Council, and the Norwegian Research Council is gratefully acknowledged.

Appendix A. Procedures and instructions for Experiment I

A.1. Procedures

Subjects are recruited to two separate rooms called room A and room B. Dictators are in room A and recipients are in room B. The subjects are welcomed and told not to talk to each other. In room B (recipients), subjects receive instructions for eliciting beliefs (i.e. guessing the outcome of the dictator game they will subsequently play). In room A (dictators), subjects receive the dictator game instructions. The subjects read the instructions. Thereafter they can ask questions individually.

In room B, subjects enter their guess of the mean allocation and their student identification number on a numbered form marked "guess." The experimenter collects the forms and removes the part with the student identification number.[26] Thereafter the experimenter puts each form into an unmarked envelope that contains six SEK 20 bills (and subjects in room B now also receive the instructions for the dictator game experiment). The experimenter brings the envelopes into room A. In room A, a monitor has been chosen among the subjects and he/she conducts the experiment and verifies that the procedures are followed as described in the instructions.[27] The monitor gives one envelope to each subject in room A. Subjects are asked to open the envelope and read the guess on the form and thereafter put the form back into the envelope. Thereafter the monitor calls one person at a time and the subject goes behind a screen. In private behind the screen, the

[24] As noted by a referee, there is a difference between facing a trustor with reported pessimistic beliefs and believing that most trustors are pessimists, so the argument is not quite identical.

[25] In relative terms, Charness and Dufwenberg's (2006) difference in beliefs between the beliefs in the Roll group and the Don't Roll group is roughly 25% larger, and has higher statistical significance. According to the one-sided Mann–Whitney test that they report, their difference is significant at the five percent level, whereas ours is significant only at the ten percent level (the two-sided p-value is 0.173).

[26] We need the student identification number to be able to pay the best guess the SEK 100 prize.

[27] The monitor receives SEK 120 in payment.

subject decides how many SEK 20 bills to leave in the envelope and how many to keep for his/her own use. The subject then seals the envelope and drops it in a box marked "Mail."

When all subjects in room A have made their decisions, the monitor brings the box marked "Mail" to room B. The monitor opens an envelope to check the number on the form marked "guess" and asks the person with this number to follow the monitor to an adjacent room. The monitor records the content of the envelope (the sum of money, if any, that is donated) and gives the contents to the person called. That person then leaves the experiment. The monitor continues until all envelopes have been opened. The experiment is then over.

A.2. Instructions – belief elicitation

The original instructions for Experiment I were in Swedish. Below is a translation of the instructions.

A.2.1. Instructions

Before we proceed with the actual experiment, we want you to guess the outcome of it. In the experiment you will be anonymously paired with another person in another room. The person in room A will decide how to split SEK 120 between him/herself and the person he/she has been paired with in room B. You will be in room B.

Every individual decision in room A will be anonymous towards both other participants and the experimenters (every person in room A will receive an envelope with six SEK 20 bills and go behind a screen and decide how many SEK 20 bills to leave in the envelope; every receiver in room B will then randomly receive one of these envelopes).

We want you to guess how much, on average, of the SEK 120 a person in room A will give to the person in room B. Write your guess and your student identification number on the form marked "guess," which was handed out together with these instructions (your guess must be between SEK 0 and SEK 120). The person among you whose guess is closest to the actual result in the experiment will win SEK 100; the winner will be notified by e-mail next week and can then collect the prize (if more than one person is equally close they will share the SEK 100).

A.3. Instructions – dictator game

The dictator game instructions were the same in both rooms.

A.3.1. Instructions

Thank you for participating in this experiment. In the experiment each of you will be paired with another person in another room. You will not be told who this other person is, neither during nor after the experiment. Excepting one person in room A who will be chosen to be a monitor, there is an equal number of persons in each room (A and B). **This is room A (B).** Every person in room A and room B has received these instructions. In the experiment every person in room A (except the monitor) will decide how to divide SEK 120 between him/herself and the person in room B with whom he/she has been paired.

The monitor will receive SEK 120. The monitor's task is to take care of the envelopes we will describe soon. Furthermore the monitor shall control and certify that the instructions we now go through were followed.

The experiment runs as follows. Unmarked envelopes corresponding to the number of participants have been put into a box in room A. All of these contain a smaller envelope and a numbered form marked "guess" (these guesses are described below). All the smaller envelopes contain six SEK 20 bills. The monitor hands out an envelope to each person in room A. When the persons in room A have gotten their envelopes they open the large envelope and take out the form marked "guess," but let the smaller envelope remain inside the large envelope. Each person silently reads the guess and thereafter puts the form back into the envelope.

The monitor then asks one person at a time in room A to come forward. The person takes the envelope and goes behind the screen in room A where no one else can see what happens.

Behind the screen, every person in room A has to decide how many bills, if any, to leave in the envelope. The person then pockets the remaining bills. The decision is up to each person in room A. No one else, including those conducting the experiment, will know what decision a particular person makes.

When the person behind the screen has made his/her decision he/she seals the smaller envelope and puts it in the larger envelope which is also sealed. The person then puts the envelope in the box marked "Mail" and leaves the room.

When all envelopes have been handed in the monitor takes the box with envelopes to room B. The monitor takes up an envelope from the box and opens the larger envelope to see the number on the form marked "guess." The person with that numbers follows the monitor to an adjacent room where the monitor opens the smaller envelope and writes down the contents and then gives the contents to the person in question who then leaves. The monitor continues until all envelopes have been opened. The experiment is then over.

On the form marked "guess" the person in room B has guessed the outcome of this experiment, that is, how much a person in room A on average will give to a person in room B. This guess was made before these instructions were handed out and without knowing that the person in room A would get to see the guess. The person in room B whose guess is closest to the actual outcome of the experiment will win SEK 100 (the winner will be notified by e-mail next week and can then collect the SEK 100 prize).

Appendix B. Procedures and instructions for Experiment II

B.1. Procedures

Subjects are welcomed and told no to talk to each other. They are randomly assigned to room A (trustors) and room B (trustees), where they receive experimental instructions. The instructions in room A describe the game, but contain no information about elicitation of beliefs. After the instructions are read and questions answered, the trustors receive two envelopes. One contains a NOK 50 bill and a green piece of paper of the same shape and color as the NOK 50 bill. The other envelope has an identification number. The trustors then decide whether to place the NOK 50 bill or the green piece of paper in the envelope with the identification number or in a personal envelope. After everyone has made their decision the envelopes are collected in a closed box by the experimenter in room A. The box is then given to an assistant waiting outside the room, who takes the box to another room (C).

The trustors are then asked what they believe is the most common amount that will be paid back to trustors that gave NOK 50 to the trustees. They receive a sheet of paper, marked with their identification number, and are instructed to circle the amount they believe will be paid back by most trustees. Subsequently, the guesses are collected by the experimenter in room A, who gives them to the assistant waiting outside the room. In room C, the assistant records investment decisions and guesses and places the guesses in the envelope with a matching identification number. If the trustor sent a green piece of paper, the paper is also put in the envelope with the guess. If the trustor sent NOK 50, the assistant puts NOK 250 in the envelope with the guess. The assistant then brings the box with envelopes to the experimenter in room B.

The experimenter in room B gives one envelope to each subject (trustee) in room B. The trustees who receive money decide the amount they will return to the trustors and place it in the envelope with the identification number. After the allocation is made, all envelopes with identification numbers are collected in a box by the experimenter in room B. The box is given to the assistant waiting outside the room, who registers the decisions in room C. Finally, the assistant gives the box with the envelopes to the experimenter in room A, who distributes the envelopes to the trustors. The experiment is then over.

B.2. Instructions – trustors' decision

The original instructions for Experiment II were in Norwegian. Below is a translation of the instructions.

B.2.1. Instructions

In this experiment there are two rooms. This is room A. Every subject in this room has a counterpart in room B. It is random who this person is. No one, not any person in either room, will at any time get to know the identity of their counterpart in the other room. It is the same number of subjects in each room.

After the instructions have been read, you will receive two envelopes. One of the envelopes has an identification number and contains NOK 50 and a green piece of paper.

You must decide whether to send the 50 note to your counterpart in room B, or keep the money for yourself.

If you decide to send the 50 note, it will be multiplied with five, which means that the person you are matched with will receive NOK 250. This person will then decide which of the following amounts he/she will return to you; $\{0, 50, 100, 150, 200, 250\}$. You send money to the person you are matched with in room B by placing the 50 note in the envelope with an identification number. In this case you should place the green piece of paper in the envelope with no identification number.

If you decide to keep the money, the person you are matched with will not receive anything and he/she will consequently not be able to return anything to you. In this case you should place the 50 note in the envelope without an identification number and the green piece of paper in the envelope with an identification number.

The envelopes with identification number will be collected in a box as soon as everyone in this room has made their decision. The box will then be handed over to a person in a neutral room. This person multiplies the donation with 5 and brings the envelopes to the other room.

The money you have at the end of the experiment is yours.

The subjects in room B will read this instruction: He or she has the same information as you.

For questions, please raise your hands.

Remember that the envelope with an identification number should contain either NOK 50 or a green piece of paper.

B.3. Instructions – trustors' beliefs

B.3.1. Instructions

Every subject in this room has now had the possibility to donate either 0 or NOK 50. The person you are matched with will receive NOK 250 if you donated NOK 50 and 0 if you decided to keep the NOK 50.

The person who receives the donation will then decide which of the following amounts he or she will transfer back to you: $\{0, 50, 100, 150, 200, 250\}$. You are now asked to answer the following question:

How much do you expect a person who receives NOK 250 will return?

Kindly indicate your choice on the form you receive. The money that subjects in room B actually returns are recorded by the experimenter, and everyone in this room that has answered the most common back-transfer receives a price of NOK 50. The prize will be paid at the end of the experiment.

The forms are collected as soon as they are completed by everyone.

B.4. Instructions – trustees

B.4.1. Instructions

In this experiment there are two rooms. This is room B. Every subject in this room has a counterpart in room A. It is random who this person is. No one, not any person in either room, will at any time get to know who their counterpart in the other room is. It is the same number of subjects in each room.

While you read this instruction, the subjects in room A face a choice that is described in this instruction (text in italic). This is the same text as the subjects in room A got.

After the instructions have been read, you will receive two envelopes. One of the envelopes has an identification number and contains NOK 50 and a green piece of paper.

You must decide whether to send the 50 note to your counterpart in room B, or keep the money for yourself.

If you decide to send the 50 note it will be multiplied with five, which means that the person you are matched with will receive NOK 250. This person will then decide which of the following amounts he/she will return to you; $\{0, 50, 100, 150, 200, 250\}$. *You send money to the person you are matched with in room B by placing the 50 note in the envelope with an identification number. In this case you should place the green piece of paper in the envelope with no identification number.*

If you decide to keep the money, the person you are matched with will not receive anything and he/she will consequently not be able to return anything to you. In this case you should place the 50 note in the envelope without an identification number and the green piece of paper in the envelope with an identification number.

The envelopes with identification number will be collected in a box as soon as everyone in this room has made their decision. The box will then be handed over to a person in a neutral room. This person multiplies the donation with 5 and brings the envelopes to the other room.

The money you have at the end of the experiment is yours.

The subjects in room B will read this instruction: He or she has the same information as you.

The subjects in this room will soon receive an envelope. **Do not open this envelope before you are instructed to.** In this envelope there is either NOK 250 or a green piece of paper. If there is a green piece of paper the person in room A that you are matched with decided to keep the NOK 50.

You will also receive a form that was completed by your counterpart in room A. On this form it is indicated what this person expects that a person who receives NOK 250 will return to the sender. After you have read the instruction you will be asked to open the envelope. If it contains NOK 250 you have to decide how much you want to return to the person you are matched with in room A. You can return any of these amounts: $\{0, 50, 100, 150, 200, 250\}$.

Place the money you decide to return in the envelope.

The money you do not place in the envelope, i.e., the money you decide to keep, should be hidden for the other subjects until the envelopes are collected. Your choice must not be revealed to anyone in this room.

If there is no money in the envelope, there is nothing for you to return.

The envelopes will be collected two minutes after you are told to open them.

When the envelopes are collected the experiment is over. Thanks for your participation.

Appendix C. Procedures and instructions for Experiment III

C.1. Procedures

Subjects are welcomed and told no to talk to each other. Subjects are randomly assigned to room A (principals) and room B (agents), and thereafter receive experimental instructions. The instruction in room A describes the game, but contains no information about elicitation of beliefs. The subjects read the instructions and can ask questions individually. After the instructions are read and questions answered, subjects in room A receive a sheet of paper with an identification number. They mark the sheet either IN or OUT. The sheets are collected by the experimenter in room A and handed over to an assistant waiting in room C.

Subjects in room A then receive new instructions and a form with their identification number. They are told to guess the percentage of subjects in room B that will roll the die. They are informed that if their guess is within a margin of 5 percentage units of the true percentage of rolls they will receive 100 NOK. The choices are collected by the experimenter in room A and handed over to the assistant waiting outside the room. The assistant gives the form with the guesses to the experimenter in room B.

After reading the instructions, the subjects in room B receive two forms, one is the form with guesses from room A on which they are informed about the beliefs of their counterpart. On the other form, marked with their identification number, they indicate whether they will ROLL THE DIE or NOT ROLL THE DIE. They are told to fold the form that contains their decision. The forms are then collected by the experimenter in room B. Everyone in room B then rolls the die and the experimenter writes down the outcome. The die is rolled by everyone in order not to reveal the individual choices made in the game to other subjects. The decision forms and the outcome of the die rolls are then given to the assistant waiting outside the room. In room C, the assistant calculates the payoff of each person, and the money is placed in envelopes which the assistant hands over to the experimenters in rooms A and B. The experimenters distribute the envelopes to the subjects and tell them that the experiment is over.

C.2. Instructions – principals' decisions

The original instructions for Experiment III were in Norwegian. This appendix contains a translation of the instructions.

C.2.1. Instructions

We want to emphasize that real money is at stake. The amount of money you earn in this experiment is paid out anonymously at the end of the session. No one, not the other subjects or the experimenters, will know your choices and how much money you earned in this experiment.

In this experiment there are two rooms. **This is room A.** All subjects in this room are paired with a subject in room B. You are player A and the person you are matched with player B. It is random who this person is. No one in room A or B will know who they are matched with.

You and your partner have two choices. You can decide IN or OUT. Your partner chooses either ROLL THE DIE or NOT ROLL THE DIE.

If you choose OUT, both you and your partner receives NOK 100 each. In that case the choice of your partner in room B has no payoff relevance.

If you choose IN your payoff depends on whether or not your partner chose to ROLL THE DIE or NOT ROLL THE DIE, and the face of the die if he/she chooses to ROLL THE DIE:

- If your partner in room B chooses NOT ROLL THE DIE he/she gets a payoff of NOK 280, while you get nothing.
- If your partner in room B chooses ROLL THE DIE, the payoff you receives depend on the face of the die. If the face shows 1 you receive nothing while your partner receives NOK 200. If the face of the die shows 2, 3, 4, 5 or 6 you receive NOK 240, while your partner receives NOK 200.

The payoffs to A and B are given in this table

	A gets	B gets
A chooses OUT	100	100
A chooses IN, B chooses NOT ROLL THE DIE	0	280
A chooses IN, B chooses ROLL THE DIE and the die shows 1	0	200
A chooses IN, B chooses ROLL THE DIE and the die shows 2, 3, 4, 5 or 6	240	200

The subjects in room B have received the same information as the subjects in room A.

Your partner in room B make their decision before they know whether A (you) choose IN or OUT. But since their choice only has an impact if A chooses IN, we ask them to make their choice as if their partner in A has decided to choose IN. The die will be rolled after B has decided either to ROLL THE DIE or NOT ROLL THE DIE.

Together with this instruction you have received a form with your identification number. Make your choice on this form. When everyone in this room has made their choice the forms will be collected by the person that distributed the instruction.

Every participant receives NOK 50 as a show up payment in addition to what you may earn in the experiment.

Please raise your hand if you have questions.

C.3. Instructions – principals' beliefs

C.3.1. Instructions

The participants in room B can choose between ROLL THE DIE and NOT ROLL THE DIE. We now ask you to guess the percentage of participants in room B that chooses to ROLL THE DIE.

Please indicate your guess on the form you soon will receive. Those that are within a margin of 5 percentage units of the percentage of B subjects that actually choose ROLL THE DIE will receive a prize of NOK 50. The prize will be paid at the end of the session.

The form is collected as soon as everyone in the room has completed it.

C.3.2. Instructions (player B (agent))

We want to emphasize that real money is at stake. The amount of money you earn in this experiment is paid out anonymously at the end of the session. No one, not the other subjects or the experimenters, will know your choices and how much money you earned in this experiment.

In this experiment there are two rooms. **This is room B.** All subjects in this room are paired with a subject in room A. You are player B and the person you are matched with is player A. It is random who this person is. No one in room A or B will know who they are matched with.

You and your partner have two choices. Your partner can choose IN or OUT. You choose either to ROLL THE DIE or NOT ROLL THE DIE.

If your partner choose OUT, both you and your partner receives NOK 100 each. In that case your choice has no impact on payoffs.

If your partner chooses IN your payoff depends on whether or not you chose to ROLL THE DIE or NOT ROLL THE DIE, and the face of the die if you choose ROLL THE DIE:

- If you choose NOT ROLL THE DIE you get NOK 280, while your partner gets nothing.
- If you choose ROLL THE DIE, the payoffs you receive depend on the face of the die. If the face shows 1 you get NOK 200 while your partner in room A gets nothing. If the face of the die shows 2, 3, 4, 5 or 6 you receive NOK 200, while your partner in room A receives NOK 240.

The payoffs to A and B are given in this table

	A gets	B gets
A chooses OUT	100	100
A chooses IN, B chooses NOT ROLL THE DIE	0	280
A chooses IN, B chooses ROLL THE DIE and the die shows 1	0	200
A chooses IN, B chooses ROLL THE DIE and the die shows 2, 3, 4, 5 or 6	240	200

The subjects in room A receive the same information as the subjects in room B.

Note that your choice of ROLL THE DIE or NOT ROLL THE DIE has no impact on the payoffs if your partner in A decided OUT. When you make a decision you are asked to choose as if your partner in room A chose IN.

You will soon receive two forms. One of the forms contains your identification number; indicate your choice on this form. The other form is completed by your partner. After your partner had decided either to choose IN or OUT he/she was asked to guess the percentage of subjects in room B that would choose ROLL THE DIE. When A wrote down their beliefs they did not know that B would receive this information. When everyone in this room has made their choices the form will be collected by the person that distributed the instructions.

After the forms have been collected, each of you will be visited by the experimenter in this room. He will give you a die and ask you to roll it. Everyone must roll the die, independent of the choice you made earlier. The reason you all have to roll the die is that this procedure guarantees that no one in room B can infer whether subjects decided to ROLL THE DIE or NOT ROLL THE DIE. If your choice was to NOT ROLL THE DIE the face of the die have no impact on the payoffs.

Every participant receives NOK 50 as a show up payment in addition to what you may earn in the experiment.

Please raise your hand if you have questions.

References

Andreoni, J., 1989. Giving with impure altruism: Applications to charity and Ricardian equivalence. J. Polit. Economy 97, 1447–1458.
Andreoni, J., 1990. Impure altruism and donations to public goods: A theory of warm-glow giving. Econ. J. 100, 464–477.
Ariely, D., Loewenstein, G., Prelec, D., 2003. Coherent arbitrariness: Stable demand curves without stable preferences. Quart. J. Econ. 118, 73–105.
Bacharach, M., Guerra, G., Zizzo, D.J., 2007. The self-fulfilling property of trust: An experimental study. Theory Dec. 63, 349–388.
Battigalli, P., Dufwenberg, M., 2007. Guilt in games. Amer. Econ. Rev., Papers and Proc. 97, 170–176.
Battigalli, P., Dufwenberg, M., 2009. Dynamic psychological games. J. Econ. Theory 141 (1), 1–35.
Becker, G., 1974. A theory of social interactions. J. Polit. Economy 82, 1063–1093.
Berg, J.E., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games Econ. Behav. 10, 122–142.
Blanco, M., Engelmann, D., Koch, A., Normann, H.-T., 2008. Belief elicitation in experiments: Is there a hedging problem? IZA Discussion Paper 3517.
Camerer, C., 2003. Behavioral Game Theory. Princeton University Press, Princeton.
Camerer, C., Loewenstein, G., Weber, M., 1989. The curse of knowledge in economic settings: An experimental analysis. J. Polit. Economy 97, 1234–1254.
Charness, G., Dufwenberg, M., 2006. Promises and partnership. Econometrica 74, 1579–1601.
Cox, J.C., 2004. How to identify trust and reciprocity. Games Econ. Behav. 46, 260–281.
Cox, J.C., Deck, C.A., 2005. On the nature of reciprocal motives. Econ. Inquiry 43, 623–635.
Dawes, R.M., 1989. Statistical criteria for establishing a truly false consensus effect. J. Exp. Soc. Psychol. 25, 1–17.
Dawes, R.M., 1990. The potential nonfalsity of the false consensus effect. In: Hogarth, R.M. (Ed.), Insights in Decision Making: A Tribute to Hillel J. Einhorn. University of Chicago Press, Chicago.
Dufwenberg, M., Gneezy, U., 2000. Measuring beliefs in an experimental lost wallet game. Games Econ. Behav. 30, 163–182.
Dufwenberg, M., Gächter, S., Hennig-Schmidt, H., 2006. The framing of games and the psychology of strategic choice. CeDEx Discussion Paper No. 2006-20.
Engelmann, D., Strobel, M., 2000. The false consensus effect disappears if representative information and monetary incentives are given. Exper. Econ. 3, 241–260.

Engelmann, D., Strobel, M., 2007. Deconstruction and reconstruction of an anomaly. Manuscript. University of Maastricht.
Geanakoplos, J., Pearce, D., Stacchetti, E., 1989. Psychological games and sequential rationality. Games Econ. Behav. 1, 60–79.
Guerra, G., Zizzo, D.J., 2004. Trust responsiveness and beliefs. J. Econ. Behav. Organ. 55, 25–30.
Hoffman, E., McCabe, K., Shachat, K., Smith, V.L., 1994. Preferences, property rights, and anonymity in bargaining games. Games Econ. Behav. 7, 346–380.
Hoffman, E., McCabe, K., Smith, V.L., 1996. Social distance and other regarding behavior in dictator games. Amer. Econ. Rev. 86, 653–660.
Madarász, K., 2008. Information projection: Model and applications. Manuscript. University of California, Berkeley.
McCabe, K.A., Rigdon, M.L., Smith, V.L., 2003. Positive reciprocity and intentions in trust games. J. Econ. Behav. Organ. 52, 267–275.
Rabin, M., 1993. Incorporating fairness into game theory and economics. Amer. Econ. Rev. 83, 1281–1302.
Ross, L., Greene, D., House, P., 1977. The false consensus effect: An egocentric bias in social perception and attribution processes. J. Exp. Soc. Psychol. 13, 279–301.
Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases. Science 185, 1124–1131.
Vanberg, C., 2008. Why do people keep their promises? An experimental test of two explanations. Econometrica 76, 1467–1480.