

IZA DP No. 3517

Belief Elicitation in Experiments: Is there a Hedging Problem?

Mariana Blanco
Dirk Engelmann
Alexander K. Koch
Hans-Theo Normann

May 2008

Belief Elicitation in Experiments: Is there a Hedging Problem?

Mariana Blanco

Royal Holloway, University of London

Dirk Engelmann

Royal Holloway, University of London

Alexander K. Koch

*Royal Holloway, University of London
and IZA*

Hans-Theo Normann

Royal Holloway, University of London

Discussion Paper No. 3517

May 2008

IZA

P.O. Box 7240

53072 Bonn

Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Belief Elicitation in Experiments: Is there a Hedging Problem?*

Belief elicitation in economics experiments usually relies on paying subjects according to the accuracy of stated beliefs in addition to payments for other decisions. Such incentives, however, allow risk-averse subjects to hedge with their stated beliefs against adverse outcomes of other decisions in the experiment. This raises two questions: (i) can we trust the existing belief elicitation results, (ii) can we avoid potential hedging confounds? Our results instill confidence regarding both issues. We propose an experimental design that eliminates hedging opportunities, and use this to test for the empirical relevance of hedging effects in the lab. We find no evidence for hedging, comparing the standard “hedging-prone” belief elicitation treatment to a “hedging-proof” design in a sequential prisoners’ dilemma game. Our findings are strengthened by the absence of hedging even in an additional non-belief elicitation treatment using a financial investment frame, where hedging arguably would be most natural.

JEL Classification: C72, C90, G11

Keywords: belief elicitation, hedging, methods, experimental economics

Corresponding author:

Alexander K. Koch
Royal Holloway, University of London
Department of Economics
Egham TW20 0EX
United Kingdom
E-mail: Alexander.Koch@rhul.ac.uk

* Financial support from the Nuffield Foundation, grant No. SGS/34070, is gratefully acknowledged.

1 Introduction

Beliefs are at the heart of the analysis of any game with strategic interaction, and behavioral models suggest intricate ways in which beliefs may come about and evolve based on introspection and past experience. Experimental economists therefore often seek to measure subjects' beliefs.

Such belief elicitation is usually incentivized, following the standard practice in experimental economics of paying subjects according to their decisions. Hence, the closer a subject's stated beliefs are to the actual distribution of actions and events, the higher his or her payoff.

This design feature, however, changes the game of interest. In the theoretical version of the game, beliefs only indirectly affect payoffs through their impact on the actions taken by players. Under incentivized belief elicitation, stated beliefs themselves become part of the payoff-relevant action space. Often this creates an opportunity for using stated beliefs to hedge against adverse outcomes in the rest of the experiment. For example, a risk-averse subject may take an action in line with her true optimistic beliefs, but then falsely state pessimistic beliefs to insure against the risk of having taken an action leading to low payoffs *ex post*. Not only might the stated beliefs deviate from the true beliefs, if subjects use them to hedge, but decisions might be biased because hedging allows subjects to choose a riskier action. As a result, neither the observed actions in an experiment nor the stated beliefs may accurately reflect the true preferences and beliefs of the subjects relevant for the underlying theoretical game.

This raises two questions: (i) can we trust the existing belief elicitation results, and (ii) can we avoid potential hedging confounds without generating new problems? Our paper addresses these questions as follows. First, we show how one can eliminate hedging opportunities with a simple twist in the experimental design. Second, we test for the empirical relevance of hedging effects in the lab, by comparing a "hedging-proof" design to a standard "hedging-prone" belief elicitation treatment in a sequential prisoners' dilemma game. Third, to strengthen our results we run an additional treatment that uses a financial investment frame, where hedging arguably would be most natural. Subjects here complete an individual decision problem that is designed to mirror the payoff structure of the hedging-prone belief elicitation treatment.

How can we avoid potential hedging confounds? The common procedure in belief elicitation experiments is that subjects receive payment for *both* their action choices in the actual game underlying the experiment and for the accuracy of their stated beliefs. The following change makes the design hedging proof: randomly pay *either* the payoff associated to the game outcome *or* for the accuracy of subjects' stated beliefs. This is similar to procedures commonly used in lottery

choice experiments, where one lottery task is randomly selected to be relevant for payoffs (e.g., Holt 1986a, Beattie and Loomes 1997), but it has not been previously recognized to be a solution for the hedging problem.¹

To address the question whether we can trust the existing belief elicitation results, we test for the empirical relevance of hedging effects in the lab by comparing a standard hedging-prone belief elicitation treatment to its hedging-proof counterpart. As a simple game is required where beliefs are relevant, we use a one-shot sequential prisoners' dilemma setting. Participants first make second-mover decisions for the case in which the first player cooperates. They are then asked for their beliefs regarding second-mover choices of the other players in their session. Finally, they make their first-mover choice. In the hedging-prone belief elicitation treatment, subjects can insure against the risk that they are exposed to when cooperating as a first mover by stating pessimistic beliefs about the second-mover choices of the other players. The hedging-proof treatment eliminates the opportunity to insure against this risk by randomly selecting, at the end of the experiment, whether the belief task or the action choices are the basis for payments.

To preview our main findings, the comparison between the two treatments shows no evidence for hedging. Subjects who use the hedging opportunity in the hedging-prone treatment should take riskier choices in the prisoner's dilemma or, provided they choose to cooperate as first mover, state less optimistic beliefs (that is, indicate a smaller probability that second movers will cooperate). In our data, we find no evidence for either of these effects.

The sequential prisoners' dilemma belief elicitation treatments use a neutral frame. As a robustness test, we therefore run an additional non-belief elicitation treatment with a financial investment decision frame, where arguably hedging should be most natural. It involves two simple individual decision tasks with payoffs that mirror those of the sequential prisoners' dilemma, and offers a similar hedging opportunity to the one in the hedging-prone belief elicitation treatment. Again we find no evidence for hedging. This strengthens the earlier results that experimental subjects do not appear to make use of hedging opportunities. Overall, our findings thus suggest that hedging bias is not a serious concern in belief elicitation experiments and thus reassures confidence in existing experiments that could be affected by a hedging bias.

We discuss the related literature in Section 2. The experimental design, procedures and results

¹Instead, researchers have addressed the hedging problem in the following ways. Some have simply ignored it, some have chosen not to pay the subjects for stating beliefs. Most commonly, researchers have chosen the incentives for the belief task to be "small" relative to the incentives regarding decisions in the game. We will discuss the advantages and disadvantages of these procedures in detail below.

of the belief elicitation experiments are presented in Section 3 and those of the financial investment frame experiment in Section 4. Section 5 concludes the paper. Instructions for the experiments are in the appendix.

2 Related Literature

We start our review of the related literature with general issues about belief elicitation. Then we deal with the hedging problem created by belief elicitation methods.

Does belief elicitation change the observed behavior that researchers seek to analyze? One conjecture is that subjects always form beliefs as part of making their choices—therefore asking them to state beliefs should not influence behavior. Evidence on this is mixed² and partly depends on whether or not the belief elicitation part is incentivized. Asking subjects to state beliefs without payoff consequences makes them more likely to play dominant strategies in Croson’s (2000) prisoner’s dilemma game, and less receptive to payoff differences in Erev, Bornstein, and Wallsten’s (1993) public-good game. However, other experiments produce no significant behavioral differences in public-good games (Gächter and Renner 2006) and asymmetric 2x2 games with a unique mixed strategy equilibrium (Rutström and Wilcox 1996).

A fundamental principle in experimental economics is to pay subjects contingent on their choices.³ Empirically, the move from hypothetical choices to incentivized choices tends to have a stronger effect than increasing the stake size (see, e.g., the surveys by Smith and Walker 1993, Beattie and Loomes 1997, and Camerer and Hogarth 1999). Related to belief elicitation, the generally held view is that such incentives reduce the amount of “noise” in the beliefs data. For this there is both direct evidence (Gächter and Renner 2006) and indirect evidence (from survey responses in Offerman, Sonnemans, and Schram 1996, p. 827). Gächter and Renner (2006) and Croson (2000) document significant changes in subjects’ behavior relative to treatments with no belief questions when stated beliefs are rewarded based on how closely they match other subjects’ actual play. Rutström and Wilcox (1996) find an effect only for players with strong payoff asymmetries between strategies. However, Wilcox and Feltovich (2000) cannot replicate Croson’s (2000) findings, and several other

²Bhatt and Camerer’s (2005) brain-imaging experiments suggest that making choices and forming beliefs within the same game involve substantially different processes. Interestingly, the areas activated in the brain do overlap for both tasks when subjects’ choices and beliefs are in equilibrium (23% of all trials), that is, beliefs about what other subjects will do are accurate and actions are best responses to own beliefs.

³Other social sciences – most notably psychology – do not regularly use incentives in experiments (for a methodological discussion see, e.g., Hertwig and Ortmann 2001).

experimental studies report no significant behavioral changes (Nyarko and Schotter 2002, Guerra and Zizzo 2004, Costa-Gomes and Weizsäcker forthcoming).

A more subtle question is if it matters *how* beliefs are elicited. The most heavily used technique for incentivized belief elicitation is the quadratic scoring rule (e.g., Bhattacharya and Pfleiderer 1985, Holt 1986b, Selten 1998, Huck and Weizsäcker 2002, Offerman et al. 2007). A voluminous literature in statistics investigates the theoretical conditions for truthful elicitation of probability beliefs, that is, addresses incentive compatibility *within* belief elicitation tasks (e.g., Murphy and Winkler 1970, Savage 1971, Allen 1987). Hurley and Shogren (2005) test experimentally whether an induced probability can indeed be recovered using elicitation mechanisms.⁴ An alternative belief elicitation method is to reward only a perfect prediction, thus asking for the mode of the beliefs (e.g., Wilcox and Feltovich 2000, Bhatt and Camerer 2005).

We now turn to the central issue of this paper, the hedging opportunities created by belief elicitation methods. Experimenters usually simply augment choice tasks with an incentivized belief elicitation task. A problem with this is that subjects have a stake in the events about which subjective probabilities are elicited.⁵ This often creates cross-task hedging opportunities that compromise the *between* task incentive compatibility of belief elicitation mechanisms. A risk-averse subject may misreport stated beliefs to insure against the event that the choice based on the true beliefs leads to a low payoff realization in the choice task, and vice versa. Put differently, when making a risky choice, the variance of the total payoffs earned from both the choice and the belief tasks can be reduced by stating pessimistic beliefs about the outcome.

How have experimenters dealt with the hedging problem? A simple solution is not to pay the belief elicitation part at all. This effectively eliminates the hedging problem, but then calls into question whether we can trust stated beliefs. As discussed above, monetary incentives in experiments do appear to matter and this has also been shown for belief elicitation tasks. A

⁴They report that stated beliefs overestimate low and underestimate high probabilities, similar to findings in calibration studies (e.g., Slovic, Fischhoff, and Lichtenstein 1977) and the judgement and biases literature in cognitive psychology (e.g., Tversky and Kahneman 1974, Gilovich, Griffin, and Kahneman 2002).

⁵In general, such stakes include any aspects of the predicted events that influence a subject's evaluation of the payoffs. Karni (1999) gives the example of a surgeon concerned about his reputation when voicing an opinion regarding the likely outcome of an operation. Theoretically, such stakes make truthful elicitation of beliefs impossible (Karni and Safra 1995). Offerman, Sonnemans, and Schram (1996) test in an experiment (where hedging is not an issue) for one such "stake" effect, namely whether subjects report beliefs biased to justify their own action. They compare stated beliefs of players compensated both for belief and choice tasks with those of "spectators", who only state beliefs, finding no significant differences.

frequently chosen design to address the hedging problem is to keep the stakes for belief elicitation “small” relative to other choice tasks.⁶ This procedure may reduce the possible hedging bias, but it cannot remove the problem entirely. Moreover, paying “small” incentives for stating beliefs is at odds with Smith’s (1982) precept that incentives should be salient. Another alternative is not to elicit beliefs and estimate beliefs based on observed actions of players and a structural econometric model. The structural estimates, however, pose identification problems (e.g., Manski 2004) and may yield belief measures that are substantially different from what stated beliefs would be (e.g., Nyarko and Schotter 2002). We conclude that none of these procedures satisfactorily deal with the hedging problem, as they all create new problems or biases.

Our “hedging-proof” experimental design offers a very simple way of eliminating hedging opportunities, by randomly paying either the payoff from playing the game or for the accuracy of stated beliefs. This design is similar to what is known as “random lottery selection procedure” (Holt 1986a) or “random problem selection procedure” (Beattie and Loomes 1997) in lottery choice experiments. To keep overall incentives per task the same in expected terms as in the “hedging-prone” treatment one can simply adjust the exchange rate (as we do in our experiments).⁷ Even though the method to pay for only one randomly selected choice among several choices is well-tested, it has not previously been recognized that this method can also be used in belief elicitation experiments to eliminate the hedging problem. Berninghaus et al. (2006) use a similar design for belief elicitation but do not address the hedging problem. Instead, they are interested in whether subjects report beliefs biased to justify their own action, similar to the question that Offerman, Sonnemans, and Schram (1996) address with their “spectator” treatment (see footnote 5).

We are aware of only two studies that have explicitly explored hedging biases in stated beliefs. Both Nyarko and Schotter (2002) and Costa-Gomes and Weizsäcker (forthcoming) conjecture, based on *indirect* evidence from their experiments, that there is not a perceivable hedging bias. Our experiments offer *direct* evidence on this, being the first to compare in an experiment a hedging-proof design to its standard hedging-prone counterpart.

⁶The following quote from Rutström and Wilcox (1996, p. 11) nicely summarizes the idea used in many papers: “The maximum payoff for each stated belief [...] is deliberately low to make belief statements less interesting in expected payoff terms than strategy choice itself, which is typical of designs like this using a scoring rule.”

⁷In any case, reduced stake sizes in expected terms do not appear to matter in lottery experiments. In their survey Beattie and Loomes (1997) report no significant “dilution effect”. See also Starmer and Sugden (1991).

3 Belief-elicitation Experiments

3.1 Design

Our design is based on the sequential prisoner’s dilemma game in Figure 1.⁸ There are two players, the first mover (FM) and the second mover (SM), who each have to choose whether to cooperate or defect ($a^k \in \{c, d\}, k \in \{FM, SM\}$). If $a^{FM} = d$, the game ends with a payoff of 10 for both FM and SM.⁹ If $a^{FM} = c$, the payoff depends on the action of SM. Following $a^{SM} = c$, payoffs are 14 for both FM and SM; following $a^{SM} = d$, the payoff is 7 for FM and 17 for SM. The experiments use a neutral frame.¹⁰

Clearly, a rational and selfish SM will always defect. There are, however, many reasons why a SM might cooperate, among which are, for example, inequity aversion (e.g., Fehr and Schmidt 1999, Bolton and Ockenfels 2000), reciprocity (e.g., Dufwenberg and Kirchsteiger 2004, Falk and Fischbacher 2006), total surplus considerations (e.g., Charness and Rabin 2002, Engelmann and Strobel 2004), or simply decision errors. A FM’s decision whether to cooperate or defect is therefore not trivial, but depends on her belief μ about the probability that she is matched with a SM who cooperates. In the game in Figure 1, $a^{FM} = c$ is a best response for a risk-neutral rational and selfish FM if and only if $\mu \geq 3/7 \approx 43$ percent.

Our belief elicitation experiments implement the above game and have a common core structure. Each session consists of ten subjects, who all complete the following sequence of tasks only once and without receiving any feedback in between tasks.

⁸The sequential prisoners’ dilemma has been studied in experiments by Clark and Sefton (2001) and Blanco, Engelmann, and Normann (2007). It shares the fundamental property with the trust or investment game (Berg, Dickhaut, and McCabe 1995) and the gift-exchange game (Fehr, Kirchsteiger, and Riedl 1993) that Pareto gains are possible, but that initiating the trade exposes the first mover to risk. In our game, there are efficiency gains (in the sense of increases in total payoff) from cooperation both at the first stage and at the second stage. The investment game has efficiency gains only at the first stage (the pie size does not increase further if the second-mover returns money), whereas the gift-exchange game (as some “trust games” in the literature) only has efficiency gains at the second stage.

⁹In their sequential prisoner’s dilemma experiments, Clark and Sefton (2001) and Blanco, Engelmann, and Normann (2007) find that 96 percent and 94 percent, respectively, of the second movers defect when the first mover defects. Given this near unanimity, we dropped this decision to simplify the experiment and implement payoffs as if the second mover defects after first-mover defection. The second-mover decision is thus conditional on the first mover choosing to cooperate.

¹⁰We relabeled players and actions as follows: FM=*A player*, SM=*B player*, FM cooperate=*IN*, FM defect=*OUT*, SM cooperate=*LEFT*, SM defect=*RIGHT*.

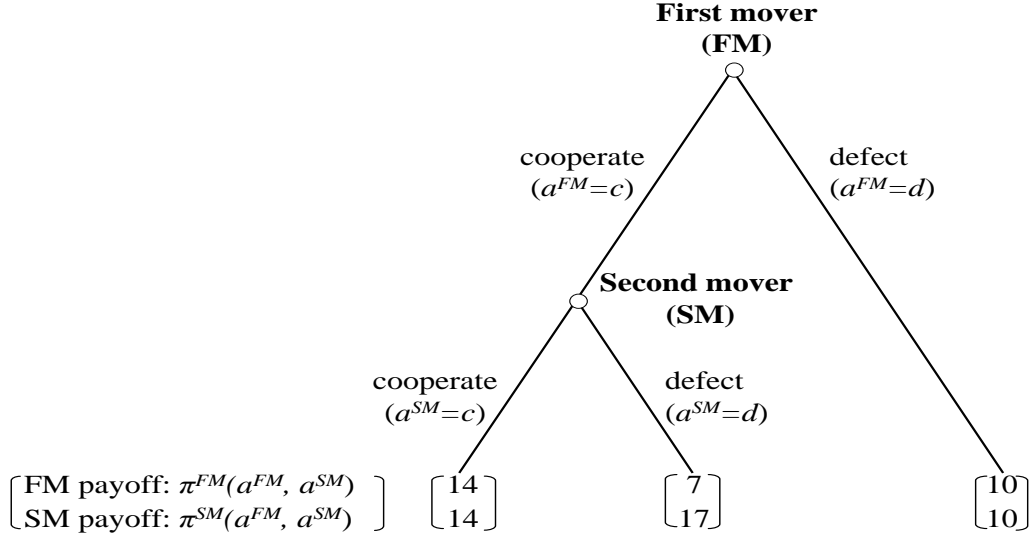


Figure 1: Sequential prisoner's dilemma game

1. *SM decision task.* Each subject makes a choice in the role of SM, for the case that FM chose to cooperate: $a_i^{SM} \in \{c, d\}$.
2. *Guess task.* Each subject is then asked to guess how many of the nine other subjects in the lab chose to cooperate in the role of SM: $g_i \in \{0, 1, \dots, 9\}$.
3. *FM decision task.* Finally, each subject makes a choice in the role of FM: $a_i^{FM} \in \{c, d\}$.

With this procedure, participants make their choices before they know whether they have the role of FM or SM. In other words, we employ the so called strategy method. We ask participants to start with making the SM choice and then elicit beliefs about the other subjects' SM choices. This makes sure that participants understand well the decision problem of the other players, about which they are making a belief statement.

Based on the profiles of choices by the ten subjects in a session, $\{a_i^{FM}, a_i^{SM}, g_i\}_{i=1, \dots, 10}$, two kinds of payoffs in experimental currency units (ECU) are computed, the decision task payoff and the guess task payoff:

Decision task payoff. The computer randomly matches all subjects in pairs. In each subject pair, one subject (say i) is randomly assigned the FM role and the other (say j) the SM role, so

		True number ^a									
		9	8	7	6	5	4	3	2	1	0
Guess ^b	9	15.0	14.8	14.3	13.3	12.0	10.4	8.3	5.9	3.1	0.0
	8	14.8	15.0	14.8	14.3	13.3	12.0	10.4	8.3	5.9	3.1
	7	14.3	14.8	15.0	14.8	14.3	13.3	12.0	10.4	8.3	5.9
	6	13.3	14.3	14.8	15.0	14.8	14.3	13.3	12.0	10.4	8.3
	5	12.0	13.3	14.3	14.8	15.0	14.8	14.3	13.3	12.0	10.4
	4	10.4	12.0	13.3	14.3	14.8	15.0	14.8	14.3	13.3	12.0
	3	8.3	10.4	12.0	13.3	14.3	14.8	15.0	14.8	14.3	13.3
	2	5.9	8.3	10.4	12.0	13.3	14.3	14.8	15.0	14.8	14.3
	1	3.1	5.9	8.3	10.4	12.0	13.3	14.3	14.8	15.0	14.8
	0	0.0	3.1	5.9	8.3	10.4	12.0	13.3	14.3	14.8	15.0

Notes: ^a the true number of the nine other subjects choosing $a^{SM} = c$; ^b the stated belief about the number of the nine other subjects choosing $a^{SM} = c$. Payoffs are based on the quadratic scoring rule in (1).

Table 1: Payoff table for the guess task (payment in experimental currency units).

that subject i 's decision task payoff is $\delta_i = \pi^{FM}(a_i^{FM}, a_j^{SM})$ and subject j 's decision task payoff is $\delta_j = \pi^{SM}(a_i^{FM}, a_j^{SM})$, as shown in Figure 1.

Guess task payoff. As is common in belief elicitation tasks, we implement a quadratic scoring rule and present it to subjects in the form of a payoff table (see Table 1). The guess task payoff γ_i depends on the accuracy of a subject's guess g_i about the true number t_i of the nine other participants in the room who have chosen to cooperate in the previous SM decision task:

$$\gamma_i = 15 \times \left[1 - \left(\frac{t_i - g_i}{9} \right)^2 \right], \quad (1)$$

where the scale parameter for adjusting the monetary payoffs is set to 15 in order to ensure that the guess and decision task payoffs are of comparable magnitude.

We have two belief-elicitation treatments. Our baseline (“hedging-prone”) treatment *Hedge* implements the procedure frequently used in belief-elicitation experiments, to base pay *both* on action choices and on the accuracy of stated beliefs. Each subject receives as final payoff the sum of decision task payoff δ_i and guess task payoff γ_i . As we will explain below, this may bias stated beliefs because of cross-task hedging.

Treatment	Final payoff	Exchange rate
<i>Hedge</i>	decision task δ_i and guess task γ_i	ECU 1 = £ 0.5
<i>NoHedge</i>	<i>either</i> decision task δ_i <i>or</i> guess task γ_i (equally likely)	ECU 1 = £ 1

Table 2: Treatments for belief elicitation experiments.

Our (“hedging-proof”) treatment *NoHedge* eliminates the cross-task hedging opportunity by basing pay *either* on the decision tasks or on the guess task. At the end of the experiment, a random draw decides whether the final payoff will be the decision task payoff δ_i or the guess task payoff γ_i . To keep incentives in each task the same across treatments in expected terms (and also to keep the total expected payoff comparable to *Hedge*), the exchange rate was doubled in treatment *NoHedge* (see Section 3.2). Table 2 summarizes the belief elicitation treatments.

Treatment *Hedge* gives subjects the opportunity to hedge with their stated beliefs against adverse outcomes of the decision task. How a subject responds to her expectations about the other players’ actions affects the correlation between her decision task payoff and her guess task payoff. The higher the stated belief in the guess task, the higher the correlation of payoffs between the guess and the decision tasks is for those who choose to cooperate as FM. In other words, a subject who states that more than half of the second movers cooperate and plays the best response to this stated belief will tend to have a high payoff in both tasks if many of the nine other players indeed cooperate as SM, and a low payoff in both tasks if many defect. The subject can, however, reduce the variance of her total payoff (at the expense of her total expected payoff) by understating in the guess task her true beliefs. This allows her to bring down or even reverse the correlation between the task payoffs. Still her payoff from FM cooperation will tend to be low if there are few cooperating second movers in the session. But the payoff from the guess task will then be higher because she stated beliefs that are more pessimistic than her true beliefs.

To address our question whether the presence of such a cross-task hedging opportunity biases stated beliefs, we compare behavior in *Hedge* with that in *NoHedge*. Because either the decision tasks or the guess task will be the basis for payments in *NoHedge*, there is no cross-task hedging benefit from stating a pessimistic belief when choosing to cooperate as first mover. *Hedge*, in contrast, allows to use stated beliefs to reduce the risk of cooperating as FM. If there is hedging in *Hedge* we should therefore observe one or both of the following two patterns:

- i) *There are more FM cooperators in Hedge than in NoHedge.*

Risk-averse players who do not hold sufficiently optimistic beliefs to make FM cooperation

their best response without the cross-task hedging opportunity might prefer to cooperate as FM when given this opportunity.

ii) *Among the FM cooperators, the stated beliefs are lower in Hedge than in NoHedge.*

Risk-averse players can reduce the risk from cooperating as FM by stating less optimistic beliefs. A strong indication of this would be if the number of subjects who cooperated as FM, even though the risk-neutral best response to their stated beliefs would be to defect, was higher in *Hedge* than in *NoHedge*. In *NoHedge* only decision errors or risk seeking behavior could explain this, while hedging provides an additional reason in *Hedge*.

The comparison with *NoHedge* allows us to assess whether subjects hedge across tasks in *Hedge* without having to measure their risk preferences.¹¹ A precise and stable measure of risk preferences would be necessary if we wanted to gauge whether subjects hedge based on the data in *Hedge* alone. For example, a participant who chooses to cooperate as FM, but states a belief that four among the other nine participants cooperate as SM, could be risk averse and hedging, or could be a risk neutral (or risk seeking) player stating her true belief. Without knowing this subject's risk preferences, there is no way of deciding whether she is hedging or not. The advantage of our design is that we can decide whether subjects are hedging or not by comparing the data across treatments. In *Hedge* the first type of players (who hedge) would behave in a systematically different way than in *NoHedge*, whereas the second type of players (who do not hedge) would not, leading to the above predictions. An additional advantage of using the comparison treatment *NoHedge* is that this permits controlling for potential influences of risk aversion on stated beliefs *within* a treatment. This issue is well known; for example, the quadratic scoring rule provides incentives for risk-averse players to state less extreme beliefs than their true beliefs (e.g., Offerman et al. 2007). Our treatment comparison nets out these effects, which are not related to our research question, because *within-task* risk reduction opportunities affect both our treatments in the same way, while cross-task hedging can occur only in *Hedge*.

¹¹Measuring risk preferences may not be conclusive because the stability of risk preferences across different tasks is not guaranteed. For example, Isaac and James (2000) elicit risk preferences in an auction and a Becker–DeGroot–Marschak mechanism with the same subjects. But they do not find within-subject stability of preferences across the two institutions. See also Friedman and Sunder (2004) and Berg, Dickhaut, and McCabe (2005).

3.2 Procedures

The experiments were carried out in the Experimental Lab of Royal Holloway, University of London, using the experimental software z-tree (Fischbacher 2007). Participants were students from various subjects, recruited through online and on-campus advertisements.

For the *Hedge* and *NoHedge* treatments, we conducted six sessions with ten participants each, providing us with thirty independent observations for those two treatments. (Because subjects make all decisions before receiving any feedback, each individual counts as an independent observation.) Thus, in total, 60 subjects participated in our belief-elicitation experiments.

At the beginning of each session, subjects were randomly assigned to separate cubicles and given time to read through the instructions.¹² Any questions were answered privately. The experiment on the computer was only started after all subjects had successfully answered a control questionnaire, requiring them to calculate simple examples on how actions determine payoffs and checking that they had understood how the final payout is determined. Prior to each task there was a short oral summary, which was delivered by the same experimenter in all sessions. At the end of the experiment, the final payout in experimental currency units (ECU) was converted into Pounds Sterling at an exchange rate of £ 0.5 per ECU (*Hedge*) and £ 1 per ECU (*NoHedge*), respectively. Sessions lasted for roughly 45 minutes with average earnings of £ 12.72 (*Hedge*: £ 12.68, *NoHedge*: £ 12.76).

3.3 Results

Before we focus on the hedging issue, we briefly summarize the main results of the experiments. Table 3 reports separately for our two treatments how many subjects cooperate/defect as FM/SM in each of the four cells ($a^{FM} = a^{SM} = c$, $a^{FM} = a^{SM} = d$, $a^{FM} = c \wedge a^{SM} = d$, $a^{FM} = d \wedge a^{SM} = c$). For each cell the table also reports in italics the average stated belief (of how many of the other nine participants in the session choose $a^{SM} = c$).

In both treatments, roughly half of the subjects cooperate as FM, and roughly half of them also cooperate as SM. The differences between treatments are negligible. Most subjects (27 of 30 in *Hedge* and 23 of 30 in *NoHedge*) make FM choices consistent with risk-neutral payoff maximization, given their stated belief and assuming standard selfish preferences. A moderate amount of risk aversion can explain the choices of all the remaining subjects. SM decisions are highly correlated with beliefs, which is consistent with a consensus effect. The term consensus effect describes the

¹²Instructions and the oral summaries mentioned below are in the appendix.

		<i>Hedge</i>			<i>NoHedge</i>			
		FM			FM			
		cooperate	defect	Σ	cooperate	defect	Σ	
SM	cooperate	# observations	13	2	15	14	3	17
		<i>average beliefs</i>	<i>6.5</i>	<i>5.5</i>		<i>6.6</i>	<i>3.0</i>	
SM	defect	# observations	3	12	15	3	10	13
		<i>average beliefs</i>	<i>4.3</i>	<i>2.2</i>		<i>4.3</i>	<i>2.6</i>	
		Σ	16	14		17	13	

Table 3: Results from treatments *Hedge* and *NoHedge*.

frequently made observation that experimental subjects tend to believe that others will behave similarly to themselves, e.g., facing a choice between two options A and B, those who choose A expect a higher rate of choices for A than those who choose B.¹³

We now turn to the main question whether or not subjects engage in cross-task hedging in the baseline treatment. We first consider the number of subjects who choose *cooperate* in the role of FM. If subjects do not hedge this number should be the same across treatments. Alternatively, hedging provides an insurance against the risk of cooperation and, thus, subjects should be more likely to cooperate as first movers in *Hedge* than in *NoHedge*. Our data yields no evidence in favor of hedging: there is no significant difference between the numbers of subjects who choose *cooperate* (Fisher’s exact test, $p = 0.999$), and the minimal difference observed (17 in *NoHedge* versus 16 in *Hedge*) actually runs counter the one predicted by the hedging hypothesis.

The second chief indicator for hedging is the belief about SM behavior stated by those who choose *cooperate* in the role of FM. Figure 2 shows the histogram of guesses of $a^{FM} = c$ subjects. If subjects do not hedge, then the beliefs stated by $a^{FM} = c$ subjects should not differ between treatments. If subjects do hedge, stated beliefs should be less optimistic in *Hedge* than in *NoHedge*.

¹³This effect has been labeled “false consensus effect” (Ross, Greene, and House 1977) in the social psychology literature and is well established there (Mullen, Atkins, Champion, Edwards, Hardly, E., and Vanderklok 1985). Dawes (1989) argues that the name is inappropriate, because the effect can only be called “false” if subjects treat their own choice differently from information about other subjects. Hence we use the term “consensus effect”, following Engelmann and Strobel (2000), who provide evidence from an experiment where beliefs are incentivized that there is a consensus effect, but no truly false consensus effect.

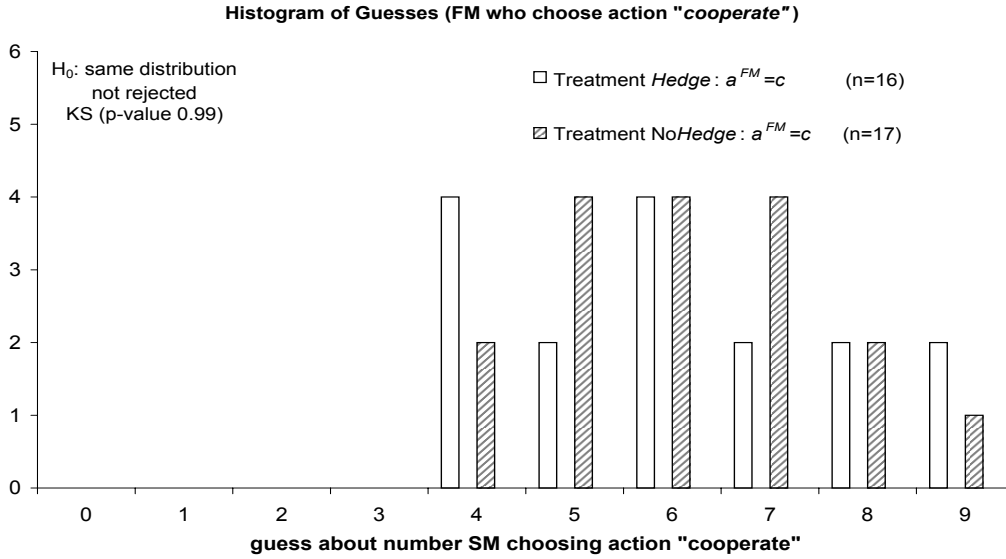


Figure 2: Histogram of beliefs about SM play stated by subjects choosing $a^{FM} = c$

As can be seen from Table 3 and Figure 2, we find no evidence for hedging. Those subjects choosing $a^{FM} = c$ hold a mean belief of 6.13 (*std. dev.* = 1.75) in *Hedge* and 6.18 (1.42) in *NoHedge*. Stated beliefs differ neither according to a Mann-Whitney U test (two-sided, $U = 131$, $p = 0.86$), nor according to a two-sample Kolmogorov-Smirnov test ($Z = 0.380$, $p = 0.999$). These findings hence are consistent with the view that subjects do not hedge.¹⁴

As noted above, most subjects play a risk-neutral best response to their stated belief: 27 out of 30 in *Hedge* and 23 out of 30 in *NoHedge*, the difference is not significant (Fisher's exact test, $p = 0.300$). If hedging actually mattered, participants in the *Hedge* treatment would bias downward stated beliefs. This could make $a^{FM} = d$ the risk-neutral best response to stated beliefs in some cases where $a^{FM} = c$ is played, and thus would predict a lower proportion of risk-neutral best responses to stated beliefs in the *Hedge* treatment. Contrary to this hypothesis, there is not a single

¹⁴We note that the beliefs of subjects who defect as first movers are also virtually identical in the two treatments, 2.64 in *Hedge* and 2.69 in *NoHedge*. This suggests that subjects are not more optimistic in *NoHedge* than in *Hedge*. Thus we rule out as explanation for the absence of a treatment difference in FM cooperation that more optimism (due to unobserved factors) in *NoHedge* pushes up FM cooperation to match that of *Hedge* and masks the impact of hedging on FM cooperation.

subject with such a pattern (those not playing a risk-neutral best response all involve a choice $a^{FM} = d$ when $a^{FM} = c$ would be the risk-neutral best response to the stated belief).

There is only one bit in the experimental data that looks like hedging at first glance. Four subjects in *Hedge* choose (cooperate, 4), but only one chooses (defect, 4). Those subjects choosing $a^{FM} = c$ could be risk averse, have a true belief of five or more and hedge risk by stating a belief of four. In *NoHedge*, where such hedging is not possible, only two subjects choose (cooperate, 4) and five choose (defect, 5). This higher share of (cooperate, 4) choices in *Hedge* could look like evidence that some players are hedging. A closer look, however, suggests that this is just the outcome of random factors unrelated to hedging. If subjects were indeed hedging we should see a decrease in the average stated belief amongst all those choosing $a^{FM} = c$ relative to corresponding the average in *NoHedge*. As discussed above, we do not find this. Moreover, comparing the rates of $a^{FM} = c$ and $a^{FM} = d$ choices among those subjects who state a belief of four, the difference between *Hedge* and *NoHedge* is far from significant (Fisher’s exact test, $p = 0.242$).

As a final piece of evidence, we note that in a post-experimental questionnaire no subject indicated cross-task hedging. Half of the subjects answered to the question “Did you think about the tasks one at a time or did you consider them jointly?” that they considered them jointly.¹⁵ However, none of them elaborated in a way suggesting that they chose the beliefs in order to hedge against the risks of FM cooperation, whereas several subjects discussed that FM cooperation is risky and said that they stated a belief that was less extreme than their true belief because this was less risky. Thus, several subjects point to within-task risk reduction, but none to across-task hedging.

4 Further Evidence Against Cross-Task Hedging: An Experiment with a Financial Frame

Our sequential prisoners’ dilemma experiment suggests that experimental subjects do not make use of cross-task hedging opportunities when they are asked for beliefs. In order to assess the robustness of this result, we conducted another experiment where the opportunity for hedging is arguably easier to recognize, and where parameters are such that incentives to hedge are strong.

¹⁵This question was aimed at revealing hedging because any subject who hedged should have been encouraged to explain this when answering the verbal comment part of this question. There are, of course, reasons other than hedging to consider the tasks jointly, most importantly to take the beliefs into account when making the FM choice.

4.1 Design and Procedures

To make hedging more natural, we additionally conducted the *Finance* experiment which has a financial investment frame that presents subjects with two individual decision problems. As explained below, the first resembles the structure of the FM choice in the sequential prisoners' dilemma experiment from Section 3, but with an exogenously given probability of SM cooperation; and the second resembles the guess task. The parameter choice is such that hedging pays even for a small degree of risk aversion.

Subjects make two decisions for which the payoffs depend on the performance of nine fictitious stocks. Each of these stocks has a probability of “performing well” of 51% and a probability of “performing poorly” of 49%. The fictitious stocks all move independently, and outcomes are based on independent draws for each subject. The first decision is whether or not to invest in one stock that is randomly drawn from the basket of nine stocks (see Figure 3). If the subject chooses not to invest in the stock (“the risk-free alternative investment”), she earns ECU 10 for sure. If she chooses to invest in the stock and it performs well, she earns ECU 14; if the stock performs poorly, she earns ECU 7. The second task is to guess how many of the nine stocks will perform well. The payoffs depend on the guess and the true number of stocks that perform well in the same way as in Table 1 for the guess task in the sequential prisoners' dilemma experiment from Section 3. The final payoff is the sum of the payoffs from both tasks, which is converted at a rate of £ 0.5 per ECU.

Note that the *Finance* experiment mirrors the structure of treatment *Hedge* for a situation where each of the other participants is known to cooperate as SM with probability 51%. However, the experiment introduces a number of differences to the *Hedge* treatment that are all aimed at making the hedging opportunity more salient. First, choices are given a financial investment frame. Second, the exogenously given probabilities mean that there are no differences in subjects' (true) beliefs here. Third, abolishing any parallel to the SM decision turns this into an individual decision experiment where strategic considerations or any other-regarding preferences cannot have an impact or distract from the hedging opportunity. Fourth, subjects first make the investment choice and then the guess, which makes the opportunity to use the belief to hedge against the risk of the investment decision more transparent.

The probabilities of 51% and 49% were chosen to make hedging pay even for a small degree of risk aversion. Given this probability, the risk-neutral (as well as risk-seeking) optimal choices are to invest in the first decision and to guess that five out of nine stocks do well. However, for a subject who chooses to invest, guessing that only four stocks will do well would reduce the variance of the

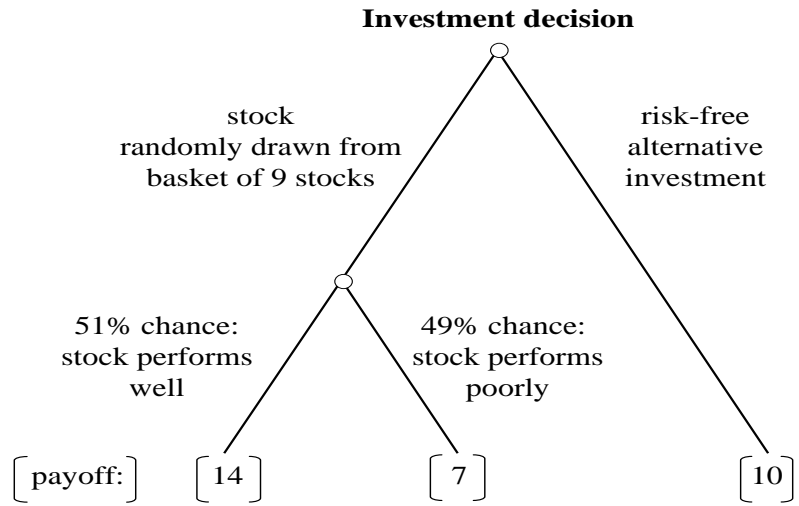


Figure 3: First investment decision in *Finance* treatment

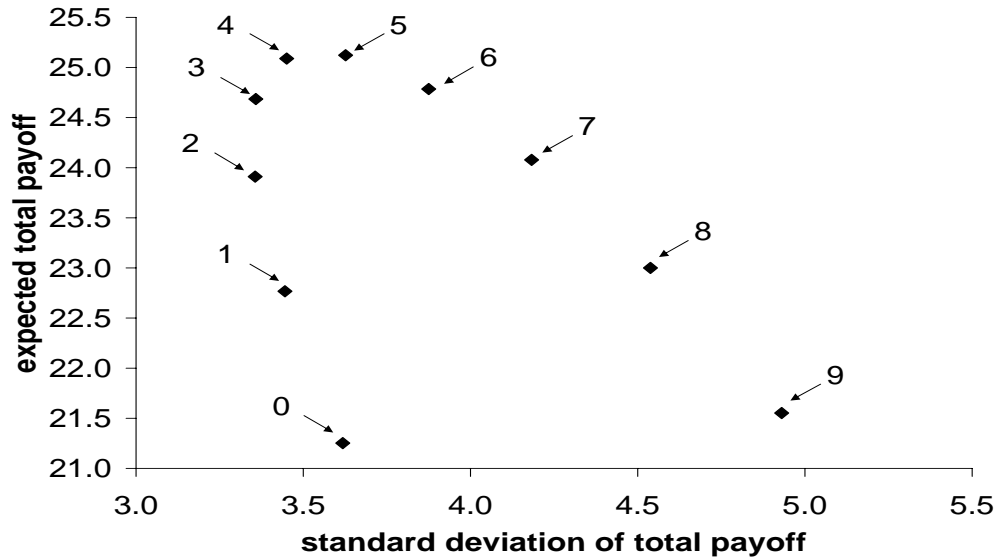


Figure 4: Total payoff (in ECU) for different guesses given that investment in stock chosen

total payoff with only a very small loss in expected payoff. Figure 4 illustrates this. Therefore, subjects who are somewhat risk-averse and understand the hedging opportunities should invest and state a belief of four (extremely risk-averse subjects should not invest and state five). Any subject investing and stating a belief of four would hence be indicative of hedging. Naturally, this might just be an error. But errors should occur in the other direction as well and also for those players who do not invest. So we can compare the rate of (invest, 4) choices with that of (invest, 6) and (not invest, 4) choices.¹⁶

For the *Finance* experiments we had 38 participants in three sessions. At the end of the experiment, the final payout in experimental currency units (ECU) was converted into Pounds Sterling at an exchange rate of £ 0.5 per ECU. Average earnings were £ 12.09. Procedures otherwise followed those for the sequential prisoners’ dilemma experiments in Section 3.2.

4.2 Results

Similar to the results for *Hedge*, we find no evidence of hedging. As reported in Table 4, among the 25 subjects who choose to invest, only two subjects guess that four of the stocks would perform well. While these two subjects could be hedging, we would rather attribute their guesses to errors. One reason is that twice as many subjects guess that six stocks will do well, which can only be driven by decision errors. Another reason is that among the 13 subjects who choose not to invest, an even higher proportion (four subjects) guess that four stocks will do well (and three subjects guess that six stocks will do well). So there is clear evidence of decision errors in the data. Intriguingly, the error rate is higher among those who decide not to invest. But still there are no “big” errors: no subject guesses less than four or more than six. As a final piece of evidence, we note that one of the two subjects who chose to invest and stated a belief of four – which would be consistent with the hedging hypothesis – noted in the post-experimental questionnaire that she considered the choices separately, whereas the other answered that she considered them jointly, without elaborating and thus without giving unambiguous indication of hedging.

To summarize, while the design of the *Finance* experiment makes hedging possibilities salient

¹⁶The argument that, if subjects do not hedge, (invest, 4) and (not invest, 4) errors should be equally frequent is based on the assumption that they consider the two decisions separately. In the hypothetical scenario that subjects first consider the most likely number of well-performing stocks and then decide whether they should invest in one stock randomly drawn from these, risk averse subjects might prefer not to invest after wrongly inferring that the best guess is four. The risk-neutral optimal choice, however, would still be to invest. Thus, if some players approach the problem in this fashion, and err that four is the best guess, some should choose to invest.

	Guess					Σ
	< 4	4	5	6	> 6	
invest	0	2	19	4	0	25
not invest	0	4	6	3	0	13

Table 4: Results from *Finance* treatment.

relative to the *Hedge* belief elicitation treatment, it still provides no evidence of subjects understanding and making use of the hedging opportunities. The fact that about a third of the subjects decide not to invest suggests that risk aversion indeed matters, and therefore the absence of hedging cannot be explained as an artefact of a risk neutral subject pool.

The absence of hedging in our experiment is related to the *under-diversification* puzzle documented in the finance literature. A prominent example is the equity home bias, reflecting the fact that individuals tend to concentrate their investments in assets from their home country and thus forgo diversification opportunities (e.g., see the survey by Lewis 1999 and, for experimental evidence, Fellner and Maciejovsky 2003). Another example is investment in own company stock. For instance, Benartzi (2001) documents that Coca-Cola employees invested up to three-quarters of their discretionary 401(k) pension contributions in Coca-Cola stock.

5 Conclusion

We presented experiments to test for a possible bias due to hedging in experiments where beliefs are elicited and incentivized. The comparison between the *Hedge* and the *NoHedge* treatments suggests no evidence whatsoever of cross-task hedging taking place. Neither do we find more first movers cooperating in *Hedge* than in *NoHedge*, nor do stated beliefs differ between treatments.

Obviously, we cannot conclude from our results that hedging is never a problem in experiments with belief elicitation (empirical non-existence proofs are impossible). Our findings do, however, support the presumption implicitly underlying previous belief elicitation experiments that cross-task hedging is not a major problem. The additional *Finance* experiment with a financial investment frame suggests that hedging is unlikely to matter even in belief elicitation experiments where hedging is more salient than in our sequential prisoners' dilemma. In the *Finance* sessions, hedging should be easier to understand and parameters are such that benefits of hedging arise even for low risk aversion — but we still find no evidence for hedging. We concede that results might differ under

diverse conditions, for example for dramatically higher incentives or for different (in particular more sophisticated) subject pools. Also, subjects may learn to exploit hedging opportunities in experiments with multiple periods if they receive feedback between periods. As the procedure we propose for making a design “hedging proof” is very simple, it seems advisable to apply it at least in such settings.

To summarize, we offer two good pieces of news for researchers who want to elicit beliefs and want to provide incentives for stating true beliefs. First, hedging does not appear to be a big worry in simple one-shot experiments. Second, for those researchers who still worry about hedging opportunities, we propose a very simple design that excludes theoretical risks of hedging across belief elicitation tasks and other tasks: simply pay randomly either for the subjects’ actions or for the stated beliefs, as in our *NoHedge* treatment.

References

- ALLEN, F. (1987): “Discovering Personal Probabilities When Utility Functions Are Unknown,” *Management Science*, 33(4), 542–544.
- BEATTIE, J., AND G. LOOMES (1997): “The Impact of Incentives upon Risky Choice Experiments,” *Journal of Risk and Uncertainty*, 14(2), 155–68.
- BENARTZI, S. (2001): “Excessive Extrapolation and the Allocation of 401(k) Accounts to Company Stock,” *Journal of Finance*, 56(5), 1747–1764.
- BERG, J., J. DICKHAUT, AND K. MCCABE (1995): “Trust, Reciprocity, and Social History,” *Games and Economic Behavior*, 10(1), 122–142.
- (2005): “Risk Preference Instability across Institutions: A Dilemma,” *Proceedings of the National Academy of Sciences USA*, 102(11), 4209–4214.
- BERNINGHAUS, S., W. GÜTH, M. V. LEVATI, AND J. QIU (2006): “Satisficing in Sales Competition: Experimental Evidence,” Discussion Papers on Strategic Interaction 2006-32, Max Planck Institute of Economics.
- BHATT, M., AND C. F. CAMERER (2005): “Self-referential Thinking and Equilibrium as States of Mind in Games: fMRI Evidence,” *Games and Economic Behavior*, 52(2), 424–459.
- BHATTACHARYA, S., AND P. PFLEIDERER (1985): “Delegated Portfolio Management,” *Journal of Economic Theory*, 36(1), 1–25.

- BLANCO, M., D. ENGELMANN, AND H.-T. NORMANN (2007): “A Within-Subject Analysis of Other-Regarding Preferences,” mimeo.
- BOLTON, G. E., AND A. OCKENFELS (2000): “ERC: A Theory of Equity, Reciprocity, and Competition,” *American Economic Review*, 90(1), 166–193.
- CAMERER, C. F., AND R. M. HOGARTH (1999): “The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework,” *Journal of Risk and Uncertainty*, 19(1-3), 7–42.
- CHARNESS, G., AND M. RABIN (2002): “Understanding Social Preferences With Simple Tests,” *The Quarterly Journal of Economics*, 117(3), 817–869.
- CLARK, K., AND M. SEFTON (2001): “The Sequential Prisoner’s Dilemma: Evidence on Reciprocation,” *Economic Journal*, 111(468), 51–68.
- COSTA-GOMES, M. A., AND G. WEIZSÄCKER (forthcoming): “Stated Beliefs and Play in Normal Form Games,” *Review of Economic Studies*.
- CROSON, R. T. A. (2000): “Thinking Like a Game Theorist: Factors Affecting the Frequency of Equilibrium Play,” *Journal of Economic Behavior & Organization*, 41(3), 299–314.
- DAWES, R. M. (1989): “Statistical Criteria for Establishing a Truly False Consensus Effect,” *Journal of Experimental Social Psychology*, 25, 1–17.
- DUFWENBERG, M., AND G. KIRCHSTEIGER (2004): “A Theory of Sequential Reciprocity,” *Games and Economic Behavior*, 47(2), 268–298.
- ENGELMANN, D., AND M. STROBEL (2000): “The False Consensus Effect Disappears if Representative Information and Monetary Incentives are Given,” *Experimental Economics*, 3(3), 241–260.
- (2004): “Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments,” *American Economic Review*, 94(4), 857–869.
- EREV, I., G. BORNSTEIN, AND T. S. WALLSTEN (1993): “The Negative Effect of Probability Assessments on Decision Quality,” *Organizational Behavior and Human Decision Processes*, 55(1), 78–94.
- FALK, A., AND U. FISCHBACHER (2006): “A Theory of Reciprocity,” *Games and Economic Behavior*, 54(2), 293–315.

- FEHR, E., G. KIRCHSTEIGER, AND A. RIEDL (1993): “Does Fairness Prevent Market Clearing? An Experimental Investigation,” *Quarterly Journal of Economics*, 108(2), 437–460.
- FEHR, E., AND K. M. SCHMIDT (1999): “A Theory Of Fairness, Competition, And Cooperation,” *Quarterly Journal of Economics*, 114(3), 817–868.
- FELLNER, G., AND B. MACIEJOVSKY (2003): “The Equity Home Bias: Contrasting an Institutional with a Behavioral Explanation,” Discussion paper, Max Planck Institute for Economics, Jena, Discussion Paper 3-2003.
- FISCHBACHER, U. (2007): “Z-Tree - Zurich Toolbox for Ready-Made Economic Experiments,” *Experimental Economics*, 10(2), 171–178.
- FRIEDMAN, D., AND S. SUNDER (2004): “Risky Curves: From Unobservable Utility to Observable Opportunity Sets,” mimeo.
- GÄCHTER, S., AND E. RENNER (2006): “The Effects of (Incentivized) Belief Elicitation in Public Good Experiments,” Discussion Papers 2006-16, The Centre for Decision Research and Experimental Economics, University of Nottingham.
- GILOVICH, T., D. GRIFFIN, AND D. KAHNEMAN (eds.) (2002): *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press, Cambridge, UK.
- GUERRA, G., AND D. J. ZIZZO (2004): “Trust Responsiveness and Beliefs,” *Journal of Economic Behavior & Organization*, 55(1), 25–30.
- HERTWIG, R., AND A. ORTMANN (2001): “Experimental Practices in Economics: A Methodological Challenge for Psychologists?” *Behavioral and Brain Sciences*, 24, 383–451.
- HOLT, C. A. (1986a): “Preference Reversals and the Independence Axiom,” *American Economic Review*, 76(3), 508–15.
- HOLT, C. A. (1986b): “Scoring-Rule Procedures for Eliciting Subjective Probability and Utility Functions,” in *Bayesian Inference and Decision Techniques*, ed. by P. Goel, and A. Zellner, pp. 279–290. Elsevier, Amsterdam.
- HUCK, S., AND G. WEIZSÄCKER (2002): “Do Players Correctly Estimate what Others Do? Evidence of Conservatism in Beliefs,” *Journal of Economic Behavior & Organization*, 47(1), 71–85.

- HURLEY, T., AND J. SHOGREN (2005): “An Experimental Comparison of Induced and Elicited Beliefs,” *Journal of Risk and Uncertainty*, 30(2), 169–188.
- ISAAC, R. M., AND D. JAMES (2000): “Just Who Are You Calling Risk Averse?” *Journal of Risk and Uncertainty*, 20(2), 177–87.
- KARNI, E. (1999): “Elicitation of Subjective Probabilities When Preferences Are State-Dependent,” *International Economic Review*, 40(2), 479–486.
- KARNI, E., AND Z. SAFRA (1995): “The Impossibility of Experimental Elicitation of Subjective Probabilities,” *Theory and Decision*, 38(3), 313–320.
- LEWIS, K. K. (1999): “Trying to Explain Home Bias in Equities and Consumption,” *Journal of Economic Literature*, 37(2), 571–608.
- MANSKI, C. F. (2004): “Measuring Expectations,” *Econometrica*, 72(5), 1329–1376.
- MULLEN, B., J. L. ATKINS, D. S. CHAMPION, C. EDWARDS, S. HARDLY, D., J. E., AND M. VANDERKLOK (1985): “The False Consensus Effect: A Meta-Analysis of 155 Hypothesis Tests,” *Journal of Experimental Social Psychology*, 21, 262–283.
- MURPHY, A. H., AND R. L. WINKLER (1970): “Scoring Rules in Probability Assessment and Evaluation,” *Acta Psychologica*, 34, 273–286.
- NYARKO, Y., AND A. SCHOTTER (2002): “An Experimental Study of Belief Learning Using Elicited Beliefs,” *Econometrica*, 70(3), 971–1005.
- OFFERMAN, T., J. SONNEMANS, AND A. SCHRAM (1996): “Value Orientations, Expectations and Voluntary Contributions in Public Goods,” *Economic Journal*, 106(437), 817–45.
- OFFERMAN, T., J. SONNEMANS, G. VAN DE KUILEN, AND P. P. WAKKER (2007): “A Truth-Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes,” Discussion paper, Econometric Institute, Erasmus University, Rotterdam.
- ROSS, L., D. GREENE, AND P. HOUSE (1977): “The False Consensus Effect: An Egocentric Bias in Social Perception and Attribution Processes,” *Journal of Experimental Social Psychology*, 13(3), 279–301.
- RUTSTRÖM, E. E., AND N. T. WILCOX (1996): “Stated Beliefs Versus Empirical Beliefs: A Methodological Inquiry and Experimental Test,” Working Paper 6-18, University of Central Florida.

- SAVAGE, L. J. (1971): “Elicitation of Personal Probabilities and Expectations,” *Journal of the American Statistical Association*, 66(336), 783–801.
- SELTEN, R. (1998): “Axiomatic Characterization of the Quadratic Scoring Rule,” *Experimental Economics*, 1(1), 43–61.
- SLOVIC, P., B. FISCHHOFF, AND S. LICHTENSTEIN (1977): “Behavioral Decision Theory,” *Annual Review of Psychology*, 28, 1–39.
- SMITH, V. L. (1982): “Microeconomic Systems as an Experimental Science,” *American Economic Review*, 72(5), 923–955.
- SMITH, V. L., AND J. M. WALKER (1993): “Monetary Rewards and Decision Cost in Experimental Economics,” *Economic Inquiry*, 31(2), 245–61.
- STARMER, C., AND R. SUGDEN (1991): “Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation,” *American Economic Review*, 81(4), 971–78.
- TVERSKY, A., AND D. KAHNEMAN (1974): “Judgment under Uncertainty: Heuristics and Biases,” *Science*, 185(4157), 1124–1131.
- WILCOX, N. T., AND N. FELTOVICH (2000): “Thinking Like a Game Theorist: Comment,” mimeo, University of Houston.

Appendix

A Instructions for the Belief Elicitation Experiments

You are now taking part in an experiment. If you read the following instructions carefully, you can, depending on your and other participants’ decisions, earn a considerable amount of money. It is therefore important that you take your time to understand the instructions. Please do not communicate with the other participants during the experiment. If you have any questions, please raise your hand and ask us. All the information you provide will be treated anonymously.

At the end of the experiment your earnings will be converted from Experimental Currency Units (ECU) to Pounds Sterling at a rate of [*Hedge*: ECU 2 = £ 1/ *NoHedge*: ECU 1 = £ 1], and paid to you in cash. Your earnings will also be treated confidentially.

SITUATION UNDERLYING THE EXPERIMENT: We start by explaining the situation underlying the experiment, which is represented in Figure 1 [corresponds to Figure 1 in this paper, but with the

neutral frame labels]. There are two people involved, Person A and Person B. Person A can choose between two options: IN or OUT. If Person A picks OUT, Person B has no choice to make, and both Person A and B get ECU 10 each. If Person A picks IN, Person B then has a choice between two options: LEFT or RIGHT. If LEFT is chosen, both Person A and B get ECU 14 each. If RIGHT is chosen, Person A gets ECU 7 and Person B gets ECU 17.

OVERVIEW OF THE EXPERIMENT: The experiment consists of three parts. You and the other participants will each make decisions both in the role of Person A and of Person B. Additionally, we will ask you to make a guess how the other participants in the room decided. At the end of the experiment, the computer will randomly assign you either the role of Person A or the role of Person B, and will also randomly match you and the other participants in pairs. Note that you will have to make your decisions without knowing the role that you will ultimately be assigned. Also, at the time when you make your decisions, you will not know the decision made by the participant matched to you. Below, we will explain how your payment from the experiment is determined. But let us first have a closer look at your tasks in the order that they will appear.

1. *Decision Task B*: You have the role of Person B in the situation described in Figure 1. Given that Person A chose IN: Do you choose LEFT or RIGHT?
2. *Guess Task*: There are 10 participants in the room, you and 9 other participants. All of them also did the above Decision Task B. How many of the 9 other participants do you think chose LEFT?
3. *Decision Task A*: Now you have the role of Person A. Do you choose IN or OUT?

PAYMENTS: [*Hedge*: At the end of the experiment you will be paid both for the Decision Tasks and for the Guess Task. Your overall payoff will be converted at a rate of $\text{ECU } 2 = \text{£ } 1$. Payoffs for the individual tasks are determined as follows.

Payoff for the Decision Tasks: As mentioned,...]

[*NoHedge*: At the end of the experiment, the computer will randomly decide whether your payment will be based on the Decision Tasks or the Guess Task. Each type of tasks is equally likely to be the one determining your payoff, and will be the same for all subjects. (This means whenever you are paid based on the Decision Tasks, also all other participants are paid based on the Decision Tasks; and whenever you are paid for the Guess Task, this is also the case for all other participants.) Your overall payoff will be converted at a rate of $\text{ECU } 1 = \text{£ } 1$. Depending on the random draw of the computer, payoffs are determined as follows.

Payoff if the random draw of the computer selects the Decision Tasks: As mentioned,] the computer will randomly and anonymously pair you with another participant in the room. One of you will randomly be assigned the role of Person A, and the other one will be assigned the role of Person B. The computer will then take your and the other participant's relevant Decision Task choices to compute your payoffs as shown in Figure 1.

[*Hedge: Payoff for the Guess Task:* In addition to the payoff for the Decision Tasks, you receive a payoff for the Guess Task, which depends...]

[*NoHedge: Payoff if the random draw of the computer selects the Guess Task:* The payoff for the Guess Task depends] on the accuracy of your guess. The better your guess, the higher will be your payoff. Take a look at Table 1 (on a separate page) [corresponds to Table 1 in this paper, but with the neutral frame labels]. The table shows how your guess and the actual choices of the other participants determine your payoff.

- You can see that a perfect guess earns you ECU 15. For example, if your guess was 6, and if there are actually 6 people who chose LEFT in Decision Task B, you get ECU 15.
- If your guess is completely off the mark you earn nothing. This occurs if you guess that 9 other participants chose LEFT, while none of them did so; or if you guess that none of the other participants chose LEFT, while all of them did so.
- Otherwise, your payoff depends on how close to accurate your guess was. For example, if 6 out of the other 9 participants chose LEFT, and your guess was that 3 participants would do so, you earn ECU 13.30.

Before starting with the actual experiment, we will ask you to answer a few control questions. Then we will go through the three parts of the experiment. There will be plenty of time before each decision to ask questions. At the end of the experiment we ask you to answer a few questions. These answers will not affect your final payment.

Are there any questions? If so, please raise your hand.

B Oral Summaries for the Belief Elicitation Experiments

Welcome to the experiment. Please do not communicate with the other participants during the experiment. If you have any questions, please raise your hand and ask us. At your seat you will find a set of instructions. Read them carefully now. Please answer the questions you find on a

separate page and raise your hand if you are finished. Before the experiment starts we will give a brief summary.

After instructions were read, before Decision Task B: To summarize: Please look at Figure 1 in the instructions. The experiment starts with Decision Task B. Next will be the Guess Task, and finally we come to Decision Task A. You will have to do each task only once. We will briefly summarize the tasks when we get to them.

[*NoHedge:* At the end of the experiment the computer will randomly decide for all participants whether the Decision Tasks are going to be the basis for payments, or the Guess task.]

At the end of the experiment, the computer randomly matches you with one of the other participants in the room. One of you will be assigned the role of Person A and the other that of Person B — both roles are equally likely. [*Hedge:* The payoffs for the Decision Tasks will then be computed based on your and the other participant's choices in the relevant Decision Tasks. In addition, the Guess task will be paid. The whole amount will then be converted to Pounds Sterling at a rate of $\text{ECU } 2 = \text{£ } 1$ / *NoHedge:* If the Decision tasks will be the basis for payments, the payoffs will then be computed based on your and the other participant's choices in the relevant Decision tasks. Otherwise, the Guess task will be paid. The payoff amount will then be converted to Pounds Sterling at a rate of $\text{ECU } 1 = \text{£ } 1$.]

We start with Decision Task B. You are asked to make a choice between LEFT and RIGHT for the case that you are assigned the role of Person B and Person A chose IN before. If you choose LEFT, both you and the other participant matched to you will get ECU 14. If you choose RIGHT, you get ECU 17 and the other participant ECU 7. Note that you will learn your actual role only at the end of the experiment. Also, if you actually are assigned the role of Person B you will learn Person A's actual choice only at the end of the experiment. [*NoHedge:* You will also learn only at the end of the experiment whether the Decision Tasks or the Guess Task will determine the payoffs.]

Are there any questions?

Between Decision Task B and Guess Task: We now come to the Guess Task. You are asked to guess how many of the 9 other participants in the room chose LEFT in the Decision Task B. Have a look at Table 1. It shows how your guess and the actual choices of the other participants determine your payoff. Also, go through the examples given in the instructions (p.3 bottom).

Are there any questions?

Between Guess Task and Decision Task A: We now come to Decision Task A. You are asked to make a choice between IN and OUT for the case that you are assigned the role of Person A. If you choose IN, your payoff and that of the other participant matched to you in the role of Person B

depend on the choice between LEFT and RIGHT of that participant, as described in Figure 1. If you choose OUT, both of you receive ECU 10, and the choice of the other participant matched to you is irrelevant for payoffs. Again, you learn your actual role only at the end of the experiment.

Are there any questions?

C Instructions for the Experiment with a Financial Frame

You are now ... treated confidentially [as in instructions for belief elicitation experiments (Appendix A), with exchange rate ECU 2 = £ 1.]

OVERVIEW OF THE EXPERIMENT: During the experiment you will make two investment decisions and your final earnings will depend on the sum of payoffs from both decisions. These decisions are related to a basket of 9 stocks from different sectors (a stock is an ownership share in a company). For each of these stocks there is a 51% probability that it will end up having a high value (i.e. the company does well), and a 49% probability that it will end up having a low value (i.e. the company does poorly). The sectors and thus the stocks' values are independent of each other. This means that the direction of the move in value for each stock is determined by a random draw that is made independently from that for the other stocks in the basket. In this experiment there is no interaction with other participants and the performance of stocks is determined separately by the computer for each participant. We now explain the first investment decision.

Investment Decision A: You choose between holding one randomly selected stock from the basket of 9 stocks described above and holding a risk-free alternative investment, unrelated to the stocks. If you choose to hold the risk-free alternative investment, your payoff from investment decision A will be ECU 10 for sure. If you choose to hold instead the stock, then there is a 51% probability that its value will be high (because this is true for each of the 9 stocks from which this stock is drawn). In this case your payoff from Investment Decision A will be ECU 14. And there is a 49% probability that the value of the stock will be low. Then your payoff from Investment Decision A will be ECU 7. That is, you have a chance of gaining ECU 4 relative to the risk-free alternative investment with slightly more than a 1 out of 2 chance and losing ECU 3 with a slightly less than 1 out of 2 chance. Figure 1 illustrates this [corresponds to Figure 3 in paper]. We now come to the second investment decision.

Investment Decision B: Investment Decision B is a “bet on the market”. Its payoff depends on the performance of the basket of 9 stocks described above (including the stock from Investment Decision A) and your guess. As explained above, for each of the 9 stocks there is a 51% probability

that it will have a high value and a 49% probability that it will have a low value. All 9 stocks move independently from each other, as the sectors they belong to are unrelated to each other. Thus, for example, if a stock in one sector has high value this does not influence the chances of the other stocks having high value.

Your payoff for Investment Decision B depends on two things: (i) your guess of the number of stocks that will have high value; (ii) the actual number of stocks having high value. Take a look at Table 1 (on a separate page) [corresponds to Table 1 in this paper, but with the finance frame labels], which shows how the payoff depends on your guess and the actual number of stocks having high value.

- You can see that a perfect match between your guess and the actual number of stocks having high value earns you ECU 15. For example, if your guess was 6, and 6 out of the 9 stocks have high value, you get ECU 15.
- You earn nothing if your guess was 9 and none of the stocks have high value, or if your guess 0 and all of the stocks have high value.
- Otherwise, your payoff depends on how close your guess is to the actual number of stocks having high value. For example, if 6 out of the 9 stocks have high value, and your guess was 3, you earn ECU 13.30.

Remember that the stock that you can invest in for Investment Decision A is randomly drawn from the basket of 9 stocks relevant for Investment Decision B. Thus, the more of the 9 stocks do well, the more likely it is that the stock in investment Decision A is among those that do well. For example, if all 9 stocks have high value, the value of the stock from Investment decision A will also be high. If all 9 stocks have low value, the value of the stock from Investment decision A will also be low.

PAYMENTS: At the end of the experiment your earnings from Investment Decision A and Investment Decision B will be added together and converted at a rate of $\text{ECU } 2 = \text{£ } 1$.

Before starting with the actual experiment... [as in instructions for belief elicitation experiments (Appendix A)].

D Oral Summary for the Experiment with a Financial Frame

[First part as in oral summaries for belief elicitation experiments (Appendix B).]

After instructions were read, before Investment Decision A: To summarize: The experiment starts with Investment Decision A. Next will be Investment Decision B. We will briefly summarize the tasks when we get to them. You will have to do each task only once. Your payoff for the experiment will be the sum of payoffs from both Investment decisions. The whole amount will then be converted to Pounds Sterling at a rate of ECU 2 = £ 1.

In this experiment there is no interaction with other participants and the performance of stocks is determined separately by the computer for each participant.

The computer will randomly determine for each stock whether it has high or low value. All 9 stocks move independently from each other, as the sectors they belong to are unrelated to each other. Thus, for example, if a stock in one sector has high value this does not influence the chances of the other stocks having high value. One of the 9 stocks is randomly chosen for Investment Decision A, to which we now turn. Please look at Figure 1 in the instructions. If you choose to hold the risk-free alternative investment, your payoff from investment decision A will be ECU 10 for sure. If you choose to hold instead the stock, then there is a 51% probability that its value will be high. In this case your payoff from Investment Decision A will be ECU 14. And there is a 49% probability that the value of the stock will be low. Then your payoff from Investment Decision A will be ECU 7. That is, you have a chance of gaining ECU 4 relative to the risk-free alternative investment with slightly more than a 1 out of 2 chance and losing ECU 3 with a slightly less than 1 out of 2 chance. Are there any questions?

Between Investment Decisions A and B: We now come to Investment Decision B. You are asked to guess how many of the 9 stocks in the basket do well. Have a look at Table 1. It shows how your guess and the actual number of stocks doing well determine your payoff. Also, go through the examples given in the instructions (p.3 bottom)

Are there any questions?