Exp Econ (2010) 13: 412–438
DOI 10.1007/s10683-010-9249-1

Belief elicitation in experiments: is there a hedging problem?

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

Received: 22 May 2008 / Accepted: 1 July 2010 / Published online: 22 July 2010

Abstract Belief-elicitation experiments usually reward accuracy of stated beliefs in addition to payments for other decisions. But this allows risk-averse subjects to hedge with their stated beliefs against adverse outcomes of the other decisions. So can we trust the existing belief-elicitation results? And can we avoid potential hedging confounds? We propose an experimental design that theoretically eliminates hedging op-

Electronic supplementary material The online version of this article (doi:10.1007/s10683-010-9249-1) contains supplementary material, which is available to authorized users.

M. Blanco
Economics Department, Universidad del Rosario, Calle 14 No. 4-80, Bogotá, Colombia
e-mail: mariana.blanco@urosario.edu.co

D. Engelmann (✉)
Department of Economics, University of Mannheim, L7, 3-5, 68131 Mannheim, Germany
e-mail: dirk.engelmann@uni-mannheim.de

D. Engelmann
Centre for Experimental Economics, University of Copenhagen, Copenhagen, Denmark

D. Engelmann
Economics Institute of the Academy of Sciences of the Czech Republic, Prague, Czech Republic

A.K. Koch
School of Economics and Management, Aarhus University, Building 1322, 8000 Aarhus C, Denmark
e-mail: akoch@econ.au.dk

H.-T. Normann
Düsseldorf Institute for Competition Economics (DICE), Düsseldorf University, 40225 Düsseldorf, Germany
e-mail: normann@dice.uni-duesseldorf.de

H.-T. Normann
Max-Planck Institute for Research on Collective Goods, Bonn, Germany

portunities. Using this design, we test for the empirical relevance of hedging effects in the lab. Our results suggest that hedging confounds are not a major problem unless hedging opportunities are very prominent. If hedging opportunities are transparent, and incentives to hedge are strong, many subjects do spot hedging opportunities and respond to them. The bias can go beyond players actually hedging themselves, because some expect others to hedge and best respond to this.

Keywords Belief elicitation · Hedging · Experimental economics · Experimental methodology

JEL Classification C72 · C90

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. Experimental economists therefore often seek to measure subjects' beliefs, allowing, for instance, to examine the source of deviations from equilibrium behavior in games and to distinguish between failure to best respond to beliefs and failure to form accurate beliefs (for example, Rey-Biel 2009).

Incentivized belief elicitation, however, changes the game of interest. Following the standard practice in experimental economics of paying subjects according to their decisions, belief elicitation is usually incentivized. Typically, the closer a subject's stated beliefs are to the actual distribution of actions and events, the higher his or her payoff. In the theoretical version of the game, beliefs only indirectly affect payoffs through their impact on the actions taken by players. But in the experiment, under incentivized belief elicitation, stated beliefs themselves become part of the payoff-relevant action space. And this often creates opportunities for subjects to use stated beliefs to hedge against adverse outcomes in the rest of the experiment.

As a simple illustration, consider a 2×2 coordination game where subjects also state a guess about what action the matched player will choose. Suppose the payoff for the action choice is x for coordination (both players choose the same action) and zero for miscoordination (one player chooses action A and the other action B). Suppose further that the payoff for a correct guess also is x. Then a subject can guarantee herself x by choosing A and guessing B, or the other way around. The alternative of choosing and guessing the same action yields a payoff of either $2x$ or zero. The former, risk-free strategy will be preferred by subjects who are sufficiently risk averse and do not hold extreme beliefs about the matched players' actions.[1]

Similar hedging incentives exist also in richer, less stylized experimental settings (a trust game, for example). A risk-averse subject may take an action in line with

[1] A person with CRRA coefficient r who thinks that with probability $p > 1/2$ the matched player will choose A, prefers the risk-free strategy over the expected-payoff maximizing combination of action A and guess A if and only if $x^{1-r} > p\,(2x)^{1-r}$. For example, with $r = 0.5$ this means that a subject will hedge if $p < 1/\sqrt{2} \approx 0.7$.

her true optimistic beliefs (to trust), but then falsely state pessimistic beliefs (a high probability to be exploited by the second mover), to insure against the risk of having taken an action that could lead to low payoffs. 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 (to trust even though the true beliefs are only moderately optimistic). As a result, neither the observed actions in an experiment nor the stated beliefs may accurately reflect the subjects' true preferences and beliefs 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? We addresses them as follows. First, we argue that one can (at least in theory) 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 running experiments where we compare a standard "hedging-prone" belief elicitation treatment with its "hedging-proof" counterpart.

1.1 How can one 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* for the accuracy of subjects' stated beliefs *or* pay the payoff associated with the game outcome.[2] Hedging then is no longer feasible, in theory, as high earnings in the tasks related to the game studied cannot compensate for low earnings in the belief elicitation task, or vice versa.

How the remedy works can easily be seen in the 2×2 coordination-game example. Suppose a subject thinks that with probability $p > 1/2$ the matched player will choose action A. With the hedging-proof design, the hedging strategy mentioned above (for example, choosing A and guessing B) never pays off. If the decision task is selected for payment, both the hedging strategy and choosing and guessing A yield the same payoff. But if the guess task is paid, the probability of a positive payout x is only $1 - p$ with the hedging strategy rather than $p > 1/2$ with the strategy to choose and guess A.

While the procedure we propose is not new to experimental economics—similar procedures are commonly used in lottery choice experiments, where one lottery task is randomly selected to be relevant for payoffs (for example, Holt 1986a, Beattie and Loomes 1997)—it has not been previously recognized to offer a remedy against the hedging problem caused by incentivized belief elicitation. To use an analogy, our contribution therefore is akin to showing that a drug commonly prescribed for one type of ailment can be used "off label" to treat a different disorder. Even though the hedging problem is well known, most experimental researchers have simply ignored it, or dealt with it by choosing the incentives for the belief task to be "small" relative to the incentives regarding decisions in the game. Below we discuss in more detail the different procedures that were previously used.

[2] To keep overall incentives per task the same in expected terms as they would have been in the corresponding hedging-prone treatment, one can simply adjust the exchange rate (as we do in our experiments).

1.2 Is the hedging problem really a problem in actual experiments?

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. In a first step, we use a one-shot sequential-move prisoners'-dilemma setting. There are two reasons for this. First, it is a simple game where beliefs are relevant. And second, the sequential prisoners' dilemma (SPD) is a prominent example of the kind of experiment where researchers have been interested in the elicitation of beliefs.[3]

Our SPD experiment works as follows. Participants start by making second-mover decisions for the case in which the first mover 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. Beliefs are rewarded with a quadratic scoring rule. The comparison of the two treatments is not indicative of hedging. Subjects who use the hedging opportunity in the hedging-prone treatment should make riskier choices in the prisoner's dilemma or, provided they choose to cooperate as first mover, state less optimistic beliefs. In our data, we find no evidence of either of these effects.

These findings seem to suggest that hedging bias is not a serious concern in similar belief elicitation experiments and thus instill confidence that existing experimental results are not affected by a hedging bias. There are, however, two caveats regarding our SPD experiment. First, hedging opportunities may not be easy to spot, and hedging could arise in other settings where these opportunities are more transparent. Second, the quantitative incentives to hedge are rather small in our experiment, when evaluated in terms of a CRRA utility function for instance, and this could be a reason for the absence of hedging.

To provide a more challenging test for the existence of a hedging bias, we ran an additional experiment with more transparent hedging opportunities and quantitatively stronger incentives for hedging. Specifically, we conducted a simple 2×2 coordination game experiment, as in the example above, where subjects are also asked to guess the choice of the player they are matched with. This setting creates a very obvious hedging opportunity if both tasks are paid. Here, we find substantial evidence of hedging. A sizable share of players state beliefs that are in contrast to their own action. These results thus show that hedging opportunities can be problematic if the design makes them rather prominent.

In a non-structured experimental questionnaire that followed the 2×2 coordination game experiment, many subjects in the hedging-prone treatment explicitly refer to the hedging logic; this strengthens our conclusion that hedging opportunities can be a problem. The analysis of the questionnaires further reveals that observed differences between the hedging-proof and the hedging-prone treatments may even understate the potential impact of hedging opportunities.

On the one hand, not all of those who understand the hedging logic actually hedge. Several of our subjects unambiguously mention the scope for hedging in their questionnaire answers, but state that they preferred a more risky, higher-yielding choice.

[3] The specific design was inspired by the fact that we were planning a study on this specific game and wanted to avoid a potential hedging bias (see Blanco et al. 2009).

Thus, if we had a setting with stronger incentives to hedge or a more risk-averse subject pool, there could be larger hedging biases than the one we observe. The second, more intriguing reason not to hedge is given by subjects who distinctly explain the hedging logic, but then write that they expected others to follow this logic, and that they therefore decided to play their (non-hedging) best response to this expectation. Recognizing the "obvious" hedging strategy made these subjects more certain in their beliefs, and thus eliminated for them the need to engage in hedging themselves.

On the other hand, there are reasons to engage in hedging-like behavior even in the theoretically hedging-proof design. In our coordination experiment, some of these choices are due to confusion, some result from randomization because subjects consider the matched player's choice virtually unpredictable and are indifferent between actions. Yet, interestingly, the post-experimental questionnaires also reveal that ambiguity (or uncertainty) aversion (for example, Ellsberg 1961) seems to drive some instances of "pseudo-hedging". A few subjects explain their hedging-like behavior with a desire to guarantee that they hold a "ticket" that can win a positive payoff in the lottery that decides at the end whether the payoffs from the decision task or from the belief-elicitation task are paid. Specifically, "pseudo-hedging" ensures that one takes the "right" decision in exactly one of the tasks, hence turning the final lottery into a lottery with a *known* winning probability of 50 percent. And, as we explain below, an ambiguity averse subject may well prefer this over the lottery with uncertain winning probability that maximizes the expected payoff.

To summarize, biased beliefs and actions through hedging can be a problem if incentives for hedging are highly transparent. In particular, our experiments show that belief-elicitation procedures then can change the way that subjects perceive the game. Some subjects play a best response against an expectation that others will hedge. Therefore, even for players who do not hedge themselves, actions and stated beliefs may be biased. Unfortunately, there appears to be no perfect cure for this. While a theoretically hedging-proof design reduces hedging-like behavior, it cannot eliminate all reasons for such behavior (for example, ambiguity aversion).

Next we discuss the related literature and then provide an overview of our experiments in Sect. 3. The experimental design, procedures, and results of the sequential prisoner's dilemma experiment are presented in Sect. 4, and those of the coordination game experiment in Sect. 5. Section 6 concludes the paper. Instructions for the experiments and more details on experiments and results are in an online appendix.

2 Related literature

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 mixed[4] 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 the prisoner's dilemma experiment of Croson (2000) and less receptive to payoff differences in the public-good experiment of Erev et al. (1993). However, other experiments produce no significant behavioral differences in public-good games (Gächter and Renner 2010, in press) and asymmetric 2×2 games with a unique mixed-strategy equilibrium (treatment *EC* in Rutström and Wilcox 2009).

A fundamental principle in experimental economics is to pay subjects contingent on their choices.[5] 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 2010, in press) and indirect evidence (from survey responses in Offerman et al. 1996, p. 827).

A more subtle question is if it matters *how* beliefs are elicited. The quadratic-scoring rule is among the most popular belief elicitation procedures (for example, Brier 1950, Bhattacharya and Pfleiderer 1985, Selten 1998, Huck and Weizsäcker 2002); but as it requires linear utility it is problematic if subjects may not be risk neutral (for example, Winkler and Murphy 1970; Savage 1971; Holt 1986b). An alternative is to reward only a perfect prediction, thus asking for the mode of the beliefs (for example, Wilcox and Feltovich 2000; Bhatt and Camerer 2005). Eliciting probability beliefs without making assumptions about risk preferences requires randomized reward functions (Allen 1987; Karni 2009; Schlag and van der Weele 2009). Other remedies are to correct the biased reports from scoring rules (Offerman et al. 2009), or to estimate structurally risk attitudes jointly with probability beliefs (for example, Andersen et al. 2009).

In marked contrast to this large literature on incentive compatibility *within* belief elicitation tasks, only few papers address the problem caused more generally by stakes in the events about which subjective probabilities are elicited (see Kadane and Winkler 1988 for an early discussion).[6] A prominent (and, as our paper shows, avoidable) case where such stakes arise is in experiments that pay both for decision outcomes and incentivize statements about related beliefs. As discussed in the introduction, such commonly used designs create hedging opportunities that compromise the *between* task incentive compatibility of an experiment.

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 calls into question whether one can trust stated beliefs. As discussed above,

[4]The brain-imaging experiment of Bhatt and Camerer (2005) suggests 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 in the cases where subjects' choices and beliefs are in equilibrium—that is, where beliefs about what other subjects will do are accurate and actions are best responses to own beliefs (23% of all trials).

[5]Other social sciences—most notably psychology—do not regularly use incentives in experiments (for a methodological discussion see, for example, Hertwig and Ortmann 2001).

[6]In general, such stakes include any aspects of the predicted events that influence a subject's evaluation of the outcomes. 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).

monetary incentives do appear to matter for belief-elicitation tasks. A frequently chosen design feature is to keep the stakes for belief elicitation "small" relative to other choice tasks. This procedure may reduce the possible hedging bias, but it cannot remove the problem entirely and is at odds with the precept that incentives should be strong enough to focus subjects' attention on the task. Another alternative is not to elicit beliefs and estimate them instead based on observed actions of players and a structural econometric model. This approach, however, poses identification problems (Manski 2004) and may yield belief measures that differ substantially from what stated beliefs would have been (for example, Nyarko and Schotter 2002).

Armantier and Treich (2009) match subjects with a different player for the task of predicting the probability of winning a two-bidder auction than for the bidding task. While reducing the scope for hedging, it does not eliminate hedging opportunities if there is correlation across individuals in the parameters governing the actions of players and if subjects are uncertain about these population parameters.[7]

Berninghaus et al. (2010, in press) use a similar design for belief elicitation as we propose, but do not explicitly address the hedging problem. Instead, they are interested in whether subjects report biased beliefs to justify their own action.

We are aware of only three studies that have explicitly explored hedging biases in stated beliefs. Both Nyarko and Schotter (2002) and Costa-Gomes and Weizsäcker (2008) conjecture, based on *indirect* evidence from their experiments, that there is no perceivable hedging bias. Haruvy et al. (2007) compare incentivized beliefs about the value of assets by subjects who trade in the asset market with stated beliefs by "observers" who do not trade and only get paid for belief accuracy. They find no significant differences between the accuracy of predictions of traders and observers. Our experiments are the first to provide a *direct* test of hedging in an experiment with belief elicitation.

3 Overview of the experiments

Table 1 summarizes our experimental treatments. In total, we had 282 subjects participating in nine different treatments.[8] In both locations (London and Copenhagen), participants were students from various subjects, recruited through online and on-campus advertisements. The recruiting in Copenhagen used ORSEE (Greiner 2004). All experiments were computerized, using the experimental software z-Tree (Fischbacher 2007). In this paper, we report in detail only on a subset of the treatments

[7] An extreme example, based on the 2×2 coordination game described in the introduction, helps illustrate the problem. Suppose a subject's prior puts probability $p > 1/2$ on all people in the population always choosing action A, and probability $1 - p$ on them all choosing action B. Hedging then still is possible: for example, choosing action A against player j and guessing that another player k will choose B. Our simple hedging-proof experimental design eliminates even such theoretical hedging opportunities.

[8] We ran an additional experiment aimed at highlighting the hedging opportunities by using an individual choice problem with a financial investment frame. The setting had the same incentive structure as the first-mover choice in *SPDHedge*, but transformed the sequential prisoners' dilemma game into an individual choice problem by fixing the probability of success (corresponding to second-mover cooperation). We found no evidence of hedging in this treatment. See Blanco et al. (2008) for details.

Table 1 Overview of experimental treatments

Sequential prisoners' dilemma (SPD) experiment [Sect. 4]		Coordination game experiment [Sect. 5]	
Royal Holloway, Univ. of London, UK (October 2007–March 2008)		LEE, Copenhagen University, Denmark (November 2008–March 2009)	
SPDHedge	$N = 30$	Hedge	$N = 40$
SPDNoHedge	$N = 30$	NoHedge	$N = 26$
		NoHedgeStrong	$N = 48$
		BinaryHedge (binary belief)[a]	$N = 26$
		BinaryNoHedge (binary belief)[a]	$N = 28$
		BinaryNoHedgeStrong (binary belief)[a]	$N = 26$
		SafeHedge (binary belief, safe option)[a]	$N = 28$

Total number of observations: 282

[a]The online appendix reports details on these treatments

conducted. For more details on the remaining coordination game treatments we refer to the online appendix. It also contains the instructions.

4 The sequential prisoner's dilemma (SPD) experiment

4.1 Design

The first experiment is based on the sequential prisoner's dilemma game in Fig. 1. There are two players, the first mover (FM) and the second mover (SM), who each have to choose whether to cooperate or defect ($a^k \in \{c, d\}, k \in \{\text{FM}, \text{SM}\}$). If $a^{\text{FM}} = d$, the game ends with a payoff of 10 for both FM and SM.[9] If $a^{\text{FM}} = c$, the payoff depends on the action of SM. Following $a^{\text{SM}} = c$, payoffs are 14 for both FM and SM; following $a^{\text{SM}} = d$, the payoff is 7 for FM and 17 for SM. The experiment uses a neutral frame. (We relabeled players and actions as follows: FM = *A player*, SM = *B player*, FM cooperate = *IN*, FM defect = *OUT*, SM cooperate = *LEFT*, SM defect = *RIGHT*.)

A rational and selfish SM will always defect in this game. But, as there are reasons why a SM might cooperate,[10] FM's decision whether to cooperate or defect depends on her belief p about the probability that she is matched with a SM who cooperates.

[9]In their sequential prisoner's dilemma experiments, Bolle and Ockenfels (1990), Clark and Sefton (2001), and Blanco et al. (2007) find that 95, 96, 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.

[10]These reasons include inequality aversion (for example, Fehr and Schmidt 1999; Bolton and Ockenfels 2000), reciprocity (for example, Dufwenberg and Kirchsteiger 2004; Falk and Fischbacher 2006), total surplus considerations (for example, Charness and Rabin 2002; Engelmann and Strobel 2004), or simply decision errors.

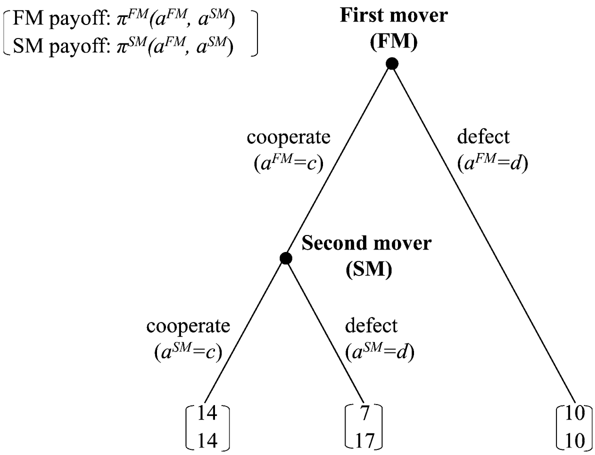


Fig. 1 Sequential prisoner's dilemma game

In the game in Fig. 1, $a^{\text{FM}} = c$ is a best response for a risk-neutral, rational, and selfish FM if and only if $p \geq 3/7$.

Our sequential prisoner's dilemma (SPD) experiment implements the above game. Each session consists of ten subjects, who all complete the following sequence of tasks only once, without receiving any feedback in between tasks. (1) *SM decision task.* Each subject makes a choice in the role of SM, for the case that FM has chosen to cooperate: $a_i^{\text{SM}} \in \{c, d\}$. (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: $g_i \in \{0, 1, \ldots, 9\}$. (3) *FM decision task.* Each subject makes a choice in the role of FM: $a_i^{\text{FM}} \in \{c, d\}$.

Note that participants make their choices before they know whether they have the role of FM or SM (strategy-elicitation method). Also, we ask participants to make the SM choice before we 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, $\{a_i^{\text{FM}}, a_i^{\text{SM}}, g_i\}_{i=1,\ldots,10}$, 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 that subject i's decision task payoff is $\delta_i = \pi^{\text{FM}}(a_i^{\text{FM}}, a_j^{\text{SM}})$ and subject j's decision task payoff is $\delta_j = \pi^{\text{SM}}(a_i^{\text{FM}}, a_j^{\text{SM}})$, as shown in Fig. 1. *Guess-task payoff:* As is common in belief-elicitation tasks, we implement a quadratic scoring rule. The guess-task payoff γ_i depends on the accuracy of a subject's guess g_i about the true number t_i of participants among the nine others in the room who have chosen to cooperate in the previous SM decision task:

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

Table 2 Treatments of the sequential prisoner's dilemma (SPD) experiment

Treatment	Final payoff	Exchange rate
SPDHedge	Decision task δ_i *and* guess task γ_i	ECU 1 = £0.5
SPDNoHedge	*either* decision task δ_i *or* guess task γ_i (equally likely)	ECU 1 = £1

where the scale parameter 15 ensures that the guess and decision task payoffs are of comparable magnitude. Guess task payoffs were rounded to multiples of ECU 0.1 and presented to subjects in the form of a payoff table (see the online appendix for the payoff table).

As discussed in Sect. 2, the quadratic-scoring rule is incentive compatible for stating the mean belief only if subjects are risk-neutral. We note, however, that our primary interest in this paper is not whether the stated beliefs are truthful, but whether decisions and beliefs are influenced by hedging opportunities. Furthermore, some of the treatments in our coordination game experiment in Sect. 5 elicit binary beliefs—where regardless of her risk preferences, a player will optimally report the state she believes to be more likely (as long as there are no hedging opportunities).

We have two treatments. In our baseline (hedging-prone) treatment *SPDHedge*, payments are based *both* on action choices and on the accuracy of stated beliefs. That is, the final payoff is the sum of decision task payoff δ_i and guess task payoff γ_i. In our (hedging-proof) treatment *SPDNoHedge*, we base payments *either* on the decision tasks *or* on the guess task. A fair random draw decides whether the final payoff is δ_i or γ_i. To keep incentives in each task the same in expected terms (and also to keep the total expected payoff comparable to *SPDHedge*), the exchange rate was doubled in treatment *SPDNoHedge*. Table 2 summarizes the sequential prisoner's dilemma treatments.

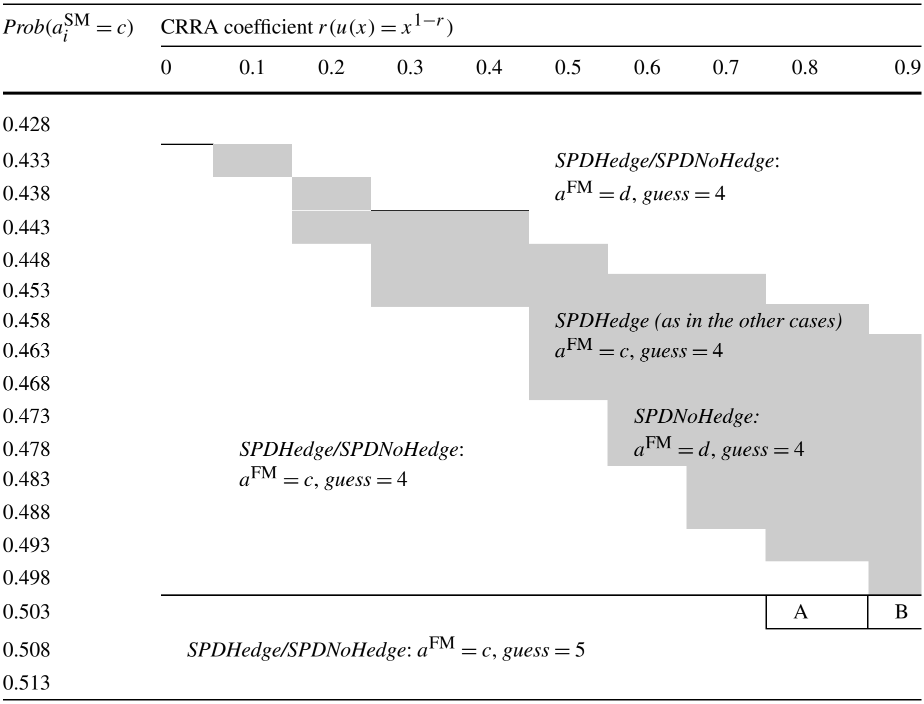
4.2 Predictions

In treatment *SPDHedge*, there is a correlation between the decision-task payoff and the guess-task payoff. 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. As both tasks are paid, the subject can reduce the variance of her total payoff by distorting her decision or her guess. For example, understating in the guess task her true beliefs allows to hedge against the risk of FM cooperation. If there are only few cooperating second movers in the session, the payoff from FM cooperation will tend to be low, but the payoff from the guess task will compensate somewhat. Since such a hedging bias can occur only in *SPDHedge*, we should observe one or both of the following two patterns:

Hypothesis 1 *There are more FM cooperators in SPDHedge than in SPDNoHedge.*

Risk-averse players, who do not hold sufficiently optimistic beliefs to make FM cooperation their best response without the hedging opportunity, might prefer to cooperate as FM when given the hedging opportunity in *SPDHedge*.

Table 3 Best responses in *SPDHedge* and *SPDNoHedge*



A: *SPDHedge:* $a^{\mathrm{FM}} = c$, *guess* $= 4$/*SPDNoHedge:* $a^{\mathrm{FM}} = c$, *guess* $= 5$

B: *SPDHedge:* $a^{\mathrm{FM}} = c$, *guess* $= 4$/*SPDNoHedge:* $a^{\mathrm{FM}} = d$, *guess* $= 5$

Hypothesis 2 *Among the FM cooperators, the stated beliefs are lower in SPDHedge than in SPDNoHedge.*

There are two reasons why such a pattern could emerge. First, a risk-averse player can reduce the risk from cooperating as FM by stating less optimistic beliefs. Second, hedging makes FM cooperation less risky, so the group of FM cooperators in *SPDHedge* may include players who cooperate even though the single-task best response to their stated beliefs would be to defect. And this would lower the average stated belief in the group of FM cooperators relative to *SPDNoHedge*.

Let us consider a specific utility function of the CRRA form to elaborate on the two hypotheses. Table 3 shows the best responses for beliefs close to the threshold required for a risk-neutral player ($r = 0$) to cooperate as first mover ($Prob(a_i^{\mathrm{SM}} = c) = 3/7 \approx 0.429$). A risk-neutral player's guess will be four for a belief up to $Prob(a_i^{\mathrm{SM}} = c) = 0.5$ and five above that level. Moving from left to right within a row (that is, holding fixed the belief) shows that a moderately risk-averse player will prefer the safer decision of FM defection if her beliefs are close to the FM-cooperation threshold of a risk-neutral player. Moving down within a column (that is, holding fixed the risk attitude) one can see at what threshold belief a risk-averse player will switch to FM cooperation.[11]

When does hedging lead to distorted FM decisions or distorted guesses? In the gray shaded area, we see that hedging allows a risk-averse player to make the risky decision of FM cooperation that she would not make in *SPDNoHedge*. So the decision is distorted but not the guess. Cell A shows that hedging may manifest itself in the form of a distorted guess, too: given risky FM cooperation, stating a more pessimistic guess provides some insurance against there being fewer SM cooperators than expected. In cell B, both the decision and the guess are distorted: stating a more pessimistic guess allows to counterbalance the riskier action of FM cooperation. For details on the calculations in Table 3 see the online appendix.

The comparison with *SPDNoHedge* allows us to assess without a measure of risk aversion whether subjects hedge in *SPDHedge*.[12] Hedging would lead to systematically different behavior in *SPDHedge* than in *SPDNoHedge*, whereas players who do not hedge would choose the same way in both treatments.

An additional advantage of using the comparison treatment *SPDNoHedge* is that it permits to control for potential influences of risk aversion on stated beliefs *within* a treatment. For example, it is well known that the quadratic scoring rule provides incentives for risk-averse players to state less extreme beliefs than their true beliefs (for example, Offerman et al. 2009). Our treatment comparison controls for 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 hedging should occur only in *SPDHedge*.

4.3 Procedures

For the *SPDHedge* and *SPDNoHedge* treatments, we conducted six sessions with ten participants each, providing us with 30 independent observations for each of those two treatments. (Because subjects make all decisions before receiving any feedback, each individual decision counts as an independent observation.)

At the beginning of each session, subjects were randomly assigned to separate cubicles and given time to read through the instructions (reproduced in the online appendix). Any questions were answered privately. The computerized experiment was only started after all subjects had successfully answered a control questionnaire. Prior to each task there was a short oral summary, which was delivered by the same

[11] Note that using CRRA to explain experimental data has several problems. For example, decision making under uncertainty by a CRRA individual depends on her current wealth level, which is almost always unobservable. Indeed, our results appear to be partly inconsistent with CRRA utility (see below). However, the CRRA model only serves to illustrate how hedging would work in this experiment, and we do not base any quantitative predictions on it.

[12] Measuring risk preferences would require yet another incentivized task, that may affect behavior in the other tasks. If we wanted to gauge whether subjects hedge based on the data in *SPDHedge* alone, a precise and stable measure of risk preferences would be necessary. The stability of risk preferences across different tasks, however, is not guaranteed. For example, Isaac and James (2000) elicit risk preferences in an auction and in a Becker-DeGroot-Marschak mechanism, without finding within-subject stability of preferences across the two institutions. See also Friedman and Sunder (2004), Berg et al. (2005).

experimenter in all sessions. Sessions lasted for roughly 45 minutes with average earnings of £ 12.72 (*SPDHedge*: £ 12.68, *SPDNoHedge*: £ 12.76). (At the time, one US dollar was approximately £ 0.5.)

4.4 Results

An overview of the behavior in the sequential prisoner's dilemma experiment is presented in the online appendix. Here we focus on the main question whether subjects engage in hedging.

We first consider the number of subjects who choose *cooperate* in the role of FM. According to Hypothesis 1, subjects should be more likely to cooperate as first movers in *SPDHedge* than in *SPDNoHedge*. If subjects do not hedge, this number should be the same across treatments. Our data yield no evidence in favor of hedging. There is no significant difference between the numbers of subjects who choose *cooperate* (two-sided Fisher exact test, $p = 0.999$; Boschloo test, $p = 0.999$),[13] and the minimal difference observed (17 in *SPDNoHedge* versus 16 in *SPDHedge*) 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 empirical cdf of guesses of $a^{\mathrm{FM}} = c$ subjects. According to Hypothesis 2, stated beliefs should be less optimistic in *SPDHedge* than in *SPDNoHedge*. As can be seen from Fig. 2, we find no evidence of this. Those subjects choosing $a^{\mathrm{FM}} = c$ stated a mean belief of 6.13 (*std. dev.* = 1.75) in *SPDHedge* and 6.18 (1.42) in *SPDNoHedge*. Neither the

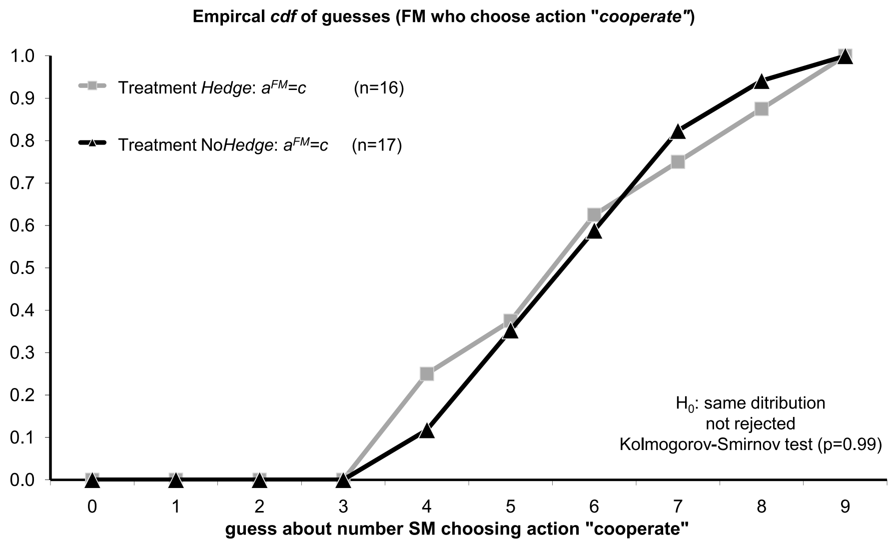


Fig. 2 Beliefs about SM play (stated by FM cooperators)

[13]Throughout the paper we report alongside the commonly used Fisher exact test also the uniformly more powerful Boschloo test. See Boschloo (1970) and the survey by Schlag 2010.

mean of the stated beliefs differ significantly (robust rank-order test, $U = -0.574$, $p = 0.566$), nor the distributions (Kolmogorov-Smirnov test, $Z = 0.380$, $p = 0.999$). Overall these findings hence suggest that subjects do not hedge.[14]

We note that most subjects play a risk-neutral best response to their stated belief: 27 out of 30 in *SPDHedge* and 23 out of 30 in *SPDNoHedge*, the difference is not significant (two-sided Fisher exact test, $p = 0.299$; Boschloo test, $p = 0.215$). If hedging actually mattered, participants in the *SPDHedge* treatment might bias downward stated beliefs. This could make $a^{\mathrm{FM}} = d$ the risk-neutral best response to stated beliefs in some cases where $a^{\mathrm{FM}} = c$ is played, and thus would predict a lower proportion of risk-neutral best responses to stated beliefs in the *SPDHedge* treatment. Contrary to this hypothesis, there is not a single subject with such a pattern.

As a final piece of evidence, we note that no subject even hinted at hedging in the post-experimental questionnaire (which is in stark contrast to the coordination game experiment reported below, where we asked the same questions). Nevertheless, around a quarter of the subjects in *SPDHedge* (8 out of 30) discussed that FM cooperation is risky and said that they had tried to reduce the payoff risk by stating a guess which was less extreme than their true belief. That is, several subjects point to within-task risk reduction, but none to hedging.

Overall, the sequential prisoner's dilemma experiment reveals no evidence of hedging. Since this represents a classic setting where researchers would want to elicit beliefs and the procedures of our hedging-prone design are also typical of such experiments, one might feel confident that worries about hedging biases in experiments are overdone.

An important counter argument against this conclusion is that the incentives to hedge in this game are quantitatively weak. A player with standard CRRA utility $u(x) = x^{1-r}$ and coefficient of risk aversion r between 0.3 and 0.5—the range commonly observed in experiments eliciting risk preferences (for example, Holt and Laury 2002)—will hedge only for a relatively narrow range of beliefs (see Table 3). So what if the hedging opportunity was more obvious? Could one be sure that subjects will not make use of this opportunity?

To follow up on this issue, the next section presents results from an experiment where the incentives for hedging are both a lot stronger (hedging is viable under CRRA preferences for a substantial range of beliefs) and a lot more transparent. Indeed, we made the hedging opportunity as transparent as we thought we could.

We wish to emphasize that it is not compelling to exonerate experiments with a potential hedging bias based on an argument that a CRRA utility function suggests only weak incentives to hedge. First, our data suggest that the CRRA utility function is not necessarily a good model of subjects' preferences. According to questionnaire responses, around a third of subjects in *SPDNoHedge* (9 out of 30) tried to reduce

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

risk by stating a less extreme guess than their true belief. Such behavior can be rationalized with CRRA utility only for relatively high degrees of risk aversion, and even then only for a tiny belief range. Second, it is conceivable that the *structure* of the tasks is what matters. If tasks suggest an easy way of risk reduction, some subjects may do so even if it is not optimal according to a CRRA utility function.

5 The coordination game experiment

5.1 Design and predictions

To provide a test of the hedging hypothesis, with stronger and more transparent incentives to hedge, we ran an additional experiment based on the simple 2×2 coordination game shown in Table 4, with belief elicitation about the action of the same matched player. Payoffs for successful coordination were chosen to be slightly asymmetric, to move subjects away from 50–50 beliefs. To avoid a clear focal point, each player has a different preferred coordination outcome. The game structure with successful coordination off the diagonal eliminates the need for potentially confusing labels for types of players to explain strategies and payoffs, because there is no difference between row and column players: "if both choose X or both Y each one earns zero; if you choose X and the other Y, your payoff is 16; if you choose Y and the other X, your payoff is 14."

Let $p = Prob(action_k = X)$ be the belief of player i about the action of the matched player k. In the game without hedging opportunities the best response is X if $p \leq 16/30 \approx 0.533$ and Y otherwise. The mixed-strategy equilibrium (which is plausible because the players are ex-ante symmetric, and this is the only symmetric equilibrium) is to play X with probability 0.533.

We report here results from treatments where incentives for stating beliefs truthfully are given in the form of a simple payoff function that is linear in the expressed certainty of the belief (increasing in the certainty if the belief is correct and decreasing if it is incorrect, see Table 5).[15] This linear function has two advantages. First, it is easy to explain and understand. Second, compared to a quadratic scoring rule, it discourages less from stating extreme beliefs and makes it easier to find evidence of hedging. If, instead of making hedging as easy as possible to detect, our goal was to objectively measure beliefs this incentive system would of course be problematic—precisely because it encourages extreme beliefs (risk neutral players should always state maximum certainty).

Table 4 Payoff table for the coordination game

Player i	Player k	
	X	Y
X	(0,0)	(16,14)
Y	(14,16)	(0,0)

[15] In the additional treatments *BinaryHedge* and *BinaryNoHedge* guesses were binary. We provide a summary of the relevant results below (for details see the online appendix).

Table 5 Payoffs for the guess task in the coordination game treatments *Hedge* and *NoHedge*

Stated belief	Actual choice of matched player *k*	
	X	*Y*
Rather *X* than Y		
x5: strongly *X*	15	0
x4	13	1
x3	11	2
x2	9	3
x1: weakly *X*	7	4
Rather *Y* than *X*		
y1: weakly *Y*	4	7
y2	3	9
y3	2	11
y4	1	13
y5: strongly *Y*	0	15

We again ran treatments with both the choice and belief tasks paid (*Hedge*) and with only one task paid (*NoHedge*). We also ran another variant of *NoHedge*. Here, an additional paragraph in the instructions stressed that it is not possible to compensate for low earnings in one task with high earnings in the other task (*NoHedgeStrong*).[16] The predictions for this treatment are obviously the same as for *NoHedge*. Parameters were chosen such that a CRRA utility function predicts hedging for reasonable degrees of risk aversion. We now summarize these predictions (see the online appendix for details).

Strength of stated beliefs: In *NoHedge*, weak belief statements should occur only for strongly risk-averse players, and then only for beliefs very close to $p = 1/2$. In *Hedge*, belief statements should always be strong (that is, they should be at the maximum certainty). The reason is that hedging can reduce risk more effectively than stating a weak belief.

Hedging: In *Hedge*, hedging should occur for beliefs ranging from around $p = 0.45$ to $p = 0.6$, if one looks at the central range of risk aversion coefficients 0.3 to 0.5 often found in experiments designed to elicit risk aversion (for example, Holt and Laury 2002). This somewhat limited band for beliefs does not seem overly problematic because it envelopes $p = 0.5$, and in our coordination game one would not expect

[16] We added the paragraph: *Remember that at the end of the experiment the computer will randomly decide whether you will be paid for the Decision Task or for the Guess Task. As only one of the tasks will be the basis for your earnings, you cannot compensate low earnings on one task with high earnings on the other task. Only what you do in the task that will actually be paid counts. So if the random draw at the end picks the Decision Task for payment, it will not help you if you made a good guess in the Guess Task. And if the random draw picks the Guess Task, it will not help you if you chose the decision for which you would have earned money in the Decision Task. You should therefore treat each task as if it was your only task and try to make the best possible decision in this task.* We note that this might create an experimenter-demand effect against hedging. But a researcher who wants to elicit beliefs without a hedging confound would exactly want to push subjects into not hedging.

subjects to hold beliefs far off from 0.5. In any case, such limits are unavoidable: for a player with a strong belief the cost of hedging (in terms of lost expected payoff) is too high to justify the insurance that it provides against a small risk.

In principle, *Hedge* offers two ways to hedge: playing X and guessing that the matched player will play X as well, or playing Y and guessing that the other will play Y as well. Because stating a weaker belief entails a penalty in terms of expected payoff, hedging should always involve extreme beliefs. That is, hedging either results in choices $(X, x5)$ or $(Y, y5)$. The former, however, dominates the latter: no matter what the other player does, it yields a higher payoff (16 instead of 15 if the matched player chooses Y, or 15 instead of 14 otherwise). But choosing the $(Y, y5)$-hedge is not implausible. A subject who thinks that others will not hedge and will be more likely to play X (for example, because they are attracted by the higher payoff) might regard hedging with a choice of $(Y, y5)$ as a "nice" move, because this makes it more likely that the counterpart will earn anything.

A final note on the predictions. As a consequence of the asymmetric payoff structure that implies a mixed-equilibrium probability different from 1/2, risk-neutral players who have beliefs $p \in (1/2, 8/15 = 0.53)$ will play X and state guess $x5$, independent of the treatment. That is, they will look as if they are hedging, even in *NoHedge*.[17] Our choice of parameters already tries to limit the belief interval where such "pseudo-hedging" is rational for risk-neutral players.[18] So the crucial comparison is whether $(X, x5)$ or $(Y, y5)$ choices are more frequent in *Hedge* than in *NoHedge* or *NoHedgeStrong*.

Hypothesis 3 *There are more* $(X, x5)$ *and* $(Y, y5)$ *choices in Hedge than in NoHedge.*

5.2 Procedures

The procedures were identical to those for the sequential prisoner's dilemma experiment (only the location of the experiments differed). We had two sessions for treatment *Hedge* with a total of 40 subjects, one session for *NoHedge* with 26 subjects, and two sessions with a total of 48 subjects in *NoHedgeStrong*. At the end of a session, the final payout in experimental currency units (ECU) was converted into Danish Kroner (DKK) at an exchange rate of DKK 7 per ECU (*Hedge*) and DKK 14 per ECU (*NoHedge* and *NoHedgeStrong*), respectively. In addition, subjects were paid a DKK 50 show-up fee. Sessions lasted for roughly 45 minutes with average earnings of DKK 144 (*Hedge*: DKK 119, *NoHedge*: DKK 148, *NoHedgeStrong*: DKK 163. At the time, one US dollar was approximately DKK 6.)

[17] In particular, the mixed-strategy equilibrium with risk-neutral players predicts that X is the most likely choice, so risk-neutral players should always guess that others will play X and mix their own choice in the game. In this equilibrium, an expected fraction 8/15 of risk-neutral subjects would look like they are hedging.

[18] For risk-averse players the belief interval where pseudo-hedging is rational in *NoHedge* is smaller. And for strong risk aversion, the stated certainty of the guess will become weaker. In contrast, in *Hedge* the belief interval where players will hedge is increasing in the degree of risk aversion, and the stated certainty of the guess should always be maximal. See the online appendix for details.

5.3 Results

Table 6 provides an overview of the key results. Belief statements contradicting own play in the coordination game—$(X, x\cdot)$ and $(Y, y\cdot)$ combinations—are slightly more frequent in *Hedge* than in the *NoHedge* treatments. There are a number of reasons why people could choose a contradicting combination; namely hedging, confusion, or beliefs that the other will play X with a probability only marginally above 50 percent. The analysis of post-experimental questionnaires will reveal another, unexpected, reason for the *NoHedge* treatments, which we will discuss as "lottery-ticket hedging" below. To identify hedging among the $(X, x\cdot)$ and $(Y, y\cdot)$ combinations, we first take a more detailed look at the stated beliefs.

Rational hedging in *Hedge* should always involve extreme stated beliefs, that is, action X coupled with guess $x5$, or action Y coupled with guess $y5$. As is shown in Table 6, such combinations are indeed twice as common in *Hedge* than in *NoHedge*. The difference in frequency is significant for *Hedge* versus the pooled *NoHedge* data (as shown in Table 7), providing first evidence of hedging.

This finding is strengthened by the analysis of post-experimental questionnaires. We classified responses by whether or not they exhibit a clear illustration of the hedging opportunity. The online appendix documents typical and particularly illuminating statements from these questionnaires. Table 6 shows that explanations of the hedging logic are surprisingly frequent in *Hedge*, whereas they are virtually absent in the *NoHedge* treatments. Table 7 reveals that the differences are significant for the comparison of *Hedge* with both the individual or the pooled *NoHedge* treatments (we report results also for the pooled data, since the *NoHedge* treatments differ very little; see also footnote 23).

Table 6 Overview of results in the coordination game experiment

	Hedge	*NoHedge*	*NoHedgeStrong*	*NoHedge* (pooled)
N	40	26	48	74
I. $(X, x\cdot)$ or $(Y, y\cdot)$[a]	16	9	17	26
(%)	40.00%	34.62%	35.42%	35.14%
II. $(X, x5)$ or $(Y, y5)$ ["(pseudo-)hedgers"]	13	4	8	12
(%)	32.50%	15.38%	16.67%	16.22%
III. Hedging possibility stated[b]	15	2	0	2
(%)	37.50%	7.69%	0.00%	2.70%
IV. $(X, x5)$ or $(Y, y5)$ and hedging poss. stated	11	0	0	0
(%)	27.50%	0.00%	0.00%	0.00%
II. as % of $(X, x\cdot)$ or $(Y, y\cdot)$ choices	81.25%	44.44%	47.06%	46.15%
IV. as % of $(X, x\cdot)$ or $(Y, y\cdot)$ choices	68.75%	0.00%	0.00%	0.00%

[a] $x\cdot$ stands for stated belief $x1$, $x2$, $x3$, $x4$, or $x5$, and $y\cdot$ for stated belief $y1$, $y2$, $y3$, $y4$, or $y5$

[b] The subject explains in the non-structured post-experimental questionnaire that there is a hedging opportunity

Table 7 Test for differences between *Hedge* and *NoHedge*

	Hedge		*NoHedge*	*NoHedgeStrong*	*NoHedge* (pooled)
"(Pseudo-)hedgers"[a]	13	vs	4	8	12
"Non-hedgers"[b]	27		22	40	62
Fisher exact test (two-sided p-value)			0.155	0.131	0.058
Boschloo test (two-sided p-value)			0.125	0.112	0.048
Hedging possibility stated[c]	15	vs	2	0	2
Not stated	25		24	48	72
Fisher exact test (two-sided p-value)			0.009	<0.001	<0.001
Boschloo test (two-sided p-value)			0.006	<0.001	<0.001

[a] $(X, x5)$ or $(Y, y5)$ choices

[b] All other choices

[c] The subject explains in the non-structured post-experimental questionnaire that there is a hedging opportunity

It seems rather striking to us that, in *Hedge*, some of the subjects who very clearly describe the hedging logic, then go on to explain why they decided not to use the hedging opportunity. A few say that they are not risk averse, and therefore did not choose to hedge.[19] Others state that they thought the hedging opportunity of playing X and guessing $x5$ was so obvious that most of the other participants would choose it; for this reason they themselves played $(Y, x5)$—the best response against $(X, x5)$.[20]

The presence of subjects who play the best response against the dominant hedging strategy suggests that looking only for a "smoking gun" of hedging may well underestimate the impact that hedging opportunities have on behavior in an experiment. In our setting, those who play such a best response to others hedging, switch from whatever they would have chosen if there was no hedging opportunity to action-guess combination $(Y, x5)$. Our experiment thus identifies a problem that has so far been neglected in discussions of problems caused by incentivized beliefs. Paying for actions as well as stated beliefs creates a completely different game. And people appear to be influenced by hedging opportunities on multiple levels, depending on their degree of sophistication. Not only may players choose differently because they hedge, but this possibility may also influence their expectation of what others do, and hence bias their choices and stated beliefs.

What about the hedging-like behavior observed in the *NoHedge* treatments? Most subjects who chose such an action-belief combination appeared to have had difficul-

[19] For example, one participant stated "…I saw the guess task as a kind of insurance of the decision task, but chose not to take the insurance because of the higher payoff with a 50% probability…"

[20] For example, one participant stated "… I thought a lot of the participants would be risk averse, and take the highest secure payoff … and I therefore gambled to win 29 ECU."

ties in finding a good strategy and made ad hoc gut decisions.[21] One contributing factor may be that the nature of our coordination game makes it difficult to have a strong belief about the matched player's behavior. Any person who thinks to have a clear idea what the other player would do, should realize that the other player might also think like this. So the logic of the mixed-strategy equilibrium kicks in (it is the only symmetric equilibrium), and the person may then realize that he or she cannot have a very certain belief about what the matched partner will do. Indeed many subjects express that they find it difficult to predict what the other will do and chose randomly in both tasks. This explains some hedging-like action-belief combinations.[22] Thus, our "therapy" against hedging is in so far successful that it eliminates clearly expressed rational hedging,[23] as well as potential biases from those who play the best response against such a "clear" hedging strategy. Our remedy does not, however, eliminate completely hedging-like combinations. But such choices might well have appeared also if we had paid nothing, or very little, for correct stated beliefs.

An interesting finding of our experiment is that not all hedging-like choices in the *NoHedge* treatments are the result of confusion or random play. We also find some players (four, and another six in further treatments summarized in the online appendix) who appear to understand the procedures well but nevertheless chose to play a hedging-like strategy. The explanation they give in the post-experimental questionnaire is that they wanted to guarantee themselves a chance of winning a positive payoff in the lottery that decides in the end whether the payoffs from the game or the guesses are paid out. One can think of them as hedging with respect to the number of "lottery tickets" they have in that lottery. These subjects appear to have a preference for having taken the "right" decision in at least one of the two tasks and thereby guaranteeing themselves a lottery ticket in the final lottery, no matter what the matched player does. Or, put differently, they have for sure a 50 percent chance of a positive payoff.[24] A trained economist will obviously call this absurd. If a subject believes that the matched player will choose X (or Y) with $p > 1/2$, then by choosing Y (X) and guessing X (Y), he will get a positive payoff with probability p (since with p he then gets a positive payoff in both tasks, which in turn leads to winning for sure a positive payoff in the final lottery). This is obviously better than winning with probability $1/2$ only (ignoring for the moment the small range of beliefs where playing X and guessing X maximizes expected payoffs).

While lottery-ticket hedging is not rationalizable with any classical utility function, it is consistent with ambiguity aversion. In our experiment, strategic ambiguity

[21] For example, one participant stated "In the decision task I picked X because it was the strongest feeling for me. The same for the guess task, X was the strongest feeling for me." and another stated "I made my choice just because my initial is Y."

[22] For example, one participant stated "I chose randomly since there is no chance of controlling the outcome."

[23] Another indication comes from *NoHedgeStrong*, which emphasizes that there is no possibility to hedge. The fact that the results in *NoHedge* and *NoHedgeStrong* are very similar indeed suggests that the hedging-like choices in the *NoHedge* treatments are not driven by subjects who mistakenly thought that such a possibility existed.

[24] For example, one participant stated "… I would rather have a 50% chance of winning either 14 or 16 than trusting my idea about the other person's behavior … "

(see for example, Eichberger et al. 2008) arises if subjects lack confidence in their probability judgment regarding other players' choices of X or Y in the decision task. This is plausible here, because subjects have no evidence to base beliefs on, neither about the same opponent nor the population in general as the game is one-shot. Even for a subject who thinks p may be substantially larger or smaller than $1/2$, this is just one of several possible conjectures. And there is uncertainty as to which conjecture is the right one.[25]

Ambiguity aversion can be captured, for example, by the maxmin expected-utility model, for which Gilboa and Schmeidler (1989) present an axiomatic foundation.[26] A player's lack of confidence in her probability judgment regarding opponents' play is reflected in a set of different possible conjectures (as opposed to a unique prior). In this model, an ambiguity-averse player will choose the action that maximizes the minimum expected payoff resulting under any of the conjectures (i.e. the payoff arising if the worst conjecture was true). For any combination of the single-task best replies for a fixed belief (choose X and guess Y or the other way around) the worst conjecture will yield an overall expected payoff of zero (the worst conjecture being that the other player chooses with probability one just that action not "bet on"). In contrast, a hedging-like combination (choose X and guess X, or choose Y and guess Y) yields a positive payoff with a probability of $1/2$, even for the worst conjecture. A sufficiently ambiguity-averse player will hence ensure to have a positive payoff in one of the two tasks no matter what the matched player does. In other words, if players indeed are ambiguity averse, then the "pseudo hedge" becomes a real hedge. For a subjective expected-utility maximizer, in contrast, such behavior would be irrational. To maximize her expected utility, she will play the single-task best response to her belief in both the decision and guess task. Ambiguity aversion is hence a plausible explanation for the "pseudo hedging", but we have no independent measure of ambiguity aversion and hence cannot verify that this is indeed underlying the observed behavior (though questionnaire statements hint at this).

5.4 Results from further treatments

In our quest to understand when hedging matters and when not, we have run a number of additional coordination game treatments, listed in Table 1. We summarize here the main findings (further details are given in the online appendix).

When we elicit beliefs in an even simpler form (players are only asked for a guess of the choice of the matched player, without any degree of certainty), we find almost no difference between the hedging-prone and hedging-proof treatments (*BinaryHedge* and *BinaryNoHedge*). The reason appears to be that in the treatments where

[25] Barberis and Thaler (2003, p. 1073) discuss as an example of ambiguity aversion a situation that nicely corresponds to our case: "a researcher might ask a subject for his estimate of the probability that a certain team will win its upcoming football match, to which the subject might respond 0.4. The researcher then asks the subject to imagine a chance machine, which will display 1 with probability 0.4 and 0 otherwise, and asks whether the subject would prefer to bet on the football game—an ambiguous bet—or on the machine, which offers no ambiguity. In general, people prefer to bet on the machine, illustrating aversion to ambiguity."

[26] We are grateful to Burkhard Schipper for extremely helpful comments on this part.

we elicit the degree of certainty together with the beliefs, we could detect hedging by focusing on the high-certainty beliefs: conscious hedging is exhibited through stated beliefs that go strongly in the direction that ensures against the risk of the action. Without the certainty information, however, these conscious hedging choices cannot be separated from random choices in line with hedging. But, similar to our earlier results, we again find in *BinaryHedge* that subjects clearly explain their hedging behavior in the post-experimental questionnaire, and that some explicitly choose a best response against the hedging strategy of others. Furthermore, in *BinaryNoHedge* there is evidence of lottery-ticket hedging.

Hedging behavior virtually disappears in *SafeHedge*—which corresponds to the *BinaryHedge* setting except that we add a 'safe guess'. A player maximizing a CRRA utility function should not make this choice. Nevertheless, lots of subjects choose this option—even though it is inferior to the also available hedging strategy. One possible explanation is that subjects stop looking for better ways of eliminating risk once they found an easy way of doing it. Indeed, while in *BinaryHedge* many subjects explain how one can hedge, few seem to spot these very same hedging opportunities when given the safe guess option.

The finding suggests that predictions made on the basis of a CRRA (or similar) utility function may have to be treated with caution. To judge whether a design is likely to cause hedging biases, it does not seem to be enough to check whether a hedging strategy could indeed maximize a reasonable utility function. The structure of the game seems to be important as well. If there is an action or belief that is obviously safe, risk-averse subjects may go for that option rather than look for possibly superior, but more complicated hedging combinations.

6 Discussion and concluding remarks

We present laboratory experiments designed to test for possible biases due to hedging opportunities in experiments where beliefs are elicited and incentivized. We do this in the context of two different classes of games. First, we use a sequential prisoner's dilemma (SPD)—as an example of a game where belief elicitation has frequently been applied and where a hedging bias could be a reason for concern. Second, we analyze a simple coordination game, where hedging incentives are quantitatively stronger and where the hedging opportunity is more obvious.

In the SPD, the comparison between our hedging-prone and hedging-proof treatments suggests no evidence whatsoever of hedging. Neither do we find more first movers cooperating in the hedging-prone treatments, nor do stated beliefs differ between the two treatments. Furthermore, not a single subject chooses what would look like the classical hedging pattern: making a risky decision and stating a pessimistic belief as an insurance against having made the wrong decision. And no subject mentions a hedging opportunity in the post-experimental questionnaire either. The absence of hedging cannot be explained by a lack of risk aversion among participants, as 43 percent of the subjects in the *Hedge* treatment stated a concern for risk reduction in the post-experimental questionnaire.

The failure to find evidence of hedging in the SPD experiment seems reassuring. It supports the presumption implicitly underlying previous belief-elicitation experiments, that hedging is not a major problem—at least in a "typical" experimental setting where hedging opportunities are not very prominent.

A potential explanation for our sequential prisoner's dilemma results, however, could be that incentives to hedge are quantitatively weak (if judged in terms of, for example, a CRRA utility function). Our second experiment thus provides a stress-test for the hypothesis that subjects do not hedge, by making hedging incentives strong and prominent.

The coordination game experiment reveals that a substantial share of subjects identify the hedging opportunity. A sizable share of the subjects choose an action-belief combination in line with hedging and express in the post-experimental questionnaire that they clearly understood the hedging incentives. A number of subjects explain how hedging works and why they themselves did not hedge. Two reasons were given for this. First, a lack of risk-aversion. Second, some subjects thought the hedging opportunity was so obvious that most others would hedge, and they therefore played a best response against this hedging strategy.

That subjects may play a best response against others using a hedging opportunity is a novel insight which, to our knowledge, has hitherto not been considered in discussions of the hedging issue. And the presence of subjects in our experiment who actually choose in this way (and refer to that logic to explain their choice) shows that the problem is not a purely academic one. The lesson is to be cautious about the way belief elicitation procedures change the game. They affect not only the riskiness of a subject's own choices for given beliefs, but they also change the strategic context—and therefore the expectations and related actions of other players who themselves may not hedge. So even if subjects' choices seem to reveal no traces of hedging, one cannot confidently conclude that there was no (indirect) hedging bias.[27] Our design may well underestimate these indirect effects of hedging opportunities. In our coordination game, the best response against the dominant hedging strategy involves playing the opposite action. So, for example, if some players hedge and a similar share play a best response against the hedging strategy, this will roughly cancel out. In a coordination game where the best response to the hedging strategy involves choosing the matching actions in the game, we would then see a more substantial shift in the pattern of actions between treatments with and without hedging opportunities.[28]

The hedging-proof treatments of the coordination game reveal that hedging-like behavior may be difficult to completely eradicate. To some extent, this may be driven

[27] Consider the following hypothetical example for our coordination game. Suppose all subjects are very sophisticated, but that they assume (most) others to be risk averse and semi-sophisticated (that is, all are "level two" in a level-of-reasoning model). In this case, all subjects would think most others are hedging, and would hence consider X to be by far the most likely choice of the player they are matched with. The best response hence is to play Y but guess $x5$. While subjects playing in this way do not hedge, the presence of a hedging opportunity nevertheless determines their behavior because it influences their beliefs. So concluding from the absence of hedging in this data that hedging opportunities did not influence behavior would be a serious mistake.

[28] We did not choose such a coordination game for the reasons discussed in Sect. 5.1, and because we expected most of our subjects to be of intermediate sophistication (that is, to possibly understand hedging but not to play the best response to others' hedging).

by the specifics of our design. It leaves many subjects uncertain about their beliefs, leading to somewhat random behavior that in some cases corresponds to an action-belief combination that suggests hedging. Some hedging-like behavior, however, also seems to be driven by ambiguity aversion.

There are a number of lessons to be learned from our experiments. First, hedging can indeed be a problem in belief-elicitation experiments. However—at least according to our results—this seems to be the case only if incentives to hedge are strong and prominent.

Second, one can take a number of simple precautions to reduce the risk of hedging. Eliciting beliefs not about the matched player, but about the whole set of other players (or set of players in the opposite role) reduces the correlation between the payoffs from the game and the payoffs from the beliefs. This already lowers the (theoretical) incentives to hedge and probably also makes the idea of hedging less prominent.[29] Our simple hedging-proof design of paying randomly either for the subjects' actions or for the stated beliefs, significantly reduces hedging in settings where such an opportunity is prominent. (As our method eliminates, at least theoretically, the hedging incentives completely, this procedure also has the nice benefit that it will satisfy the theorists in your seminars.) An alternative solution could be not to pay for beliefs at all. In many cases this may work well, because subjects have no reasons not to report their beliefs truthfully. In social-dilemma experiments, however, stated beliefs may well be used as a justification of selfish behavior—making it difficult to distinguish whether true beliefs are correlated with subjects' own behavior or whether the non-incentivized stated beliefs were biased in order to justify subjects' own behavior. Moreover, subjects might play a hedge-like strategy simply because they want to be right at least some of the time. Not paying for beliefs will not eliminate such psychological hedging.

As a third and final lesson, when attempting to eliminate hedging confounds, it may be useful to clearly explain to subjects why they should not hedge, in order to avoid confusion-driven pseudo-hedging. Furthermore, it can be useful to ask in a post-experimental questionnaire about the belief formation and how (stated) beliefs and actions interacted. These answers can reveal how an opportunity to hedge may have influenced behavior in ways not detectable in the choice data alone. If one applies our hedging-proof design, they can also be useful to distinguish ambiguity-aversion-driven pseudo-hedging from confusion. While a bias induced by ambiguity aversion seems hard to eradicate, evidence of it is still a useful warning signal. If many subjects perceive the situation as one of ambiguity, one should consider whether it is at all reasonable to elicit a measure of subjective expectations in this setting—or at least treat such measures with caution.

Acknowledgements Financial support from the Nuffield Foundation, grant No. SGS/34070, is gratefully acknowledged. We thank two anonymous referees, Tim Cason, Glenn Harrison, Burkhard Schipper,

[29] Armantier and Treich (2009) and Fehr et al. (2008) go even further and pay subjects for their belief statement based on how accurately it reflects the choice by a player other than the one they are matched with for the game. While this further reduces the correlation between game payoffs and guess payoffs, it still does not eliminate theoretical hedging opportunities as discussed in Sect. 2.

Karl Schlag and Nicholas Treich for helpful comments and suggestions. We thank the Centre for Experimental Economics (CEE) at the University of Copenhagen for giving us access to the experimental laboratory (LEE) for some of our sessions. In particular we are grateful to Jean-Robert Tyran and Marco Piovesan for their help and support and to Pablo Torija for excellent assistance. Dirk Engelmann acknowledges financial support from the institutional research grant AV0Z70850503 of the Economics Institute of the Academy of Sciences of the Czech Republic, v.v.i. Mariana Blanco thanks the Overseas Research Student Awards Scheme for financial support. Substantial parts of this research were conducted while all authors were at Royal Holloway, University of London.

References

Allen, F. (1987). Discovering personal probabilities when utility functions are unknown. *Management Science*, *33*(4), 542–544.

Andersen, S., Fountain, J., Harrison, G. W., & Rutström, E. E. (2009). Estimating subjective probabilities (Working Paper 09-01). University of Central Florida.

Armantier, O., & Treich, N. (2009). Subjective probabilities in games: an application to the overbidding puzzle. *International Economic Review*, *50*(4), 1079–1102.

Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. In G. Constantinides, M. Harris, & R. M. Stulz (Eds.), *Handbook of the economics of finance* (Vol. 1, pp. 1053–1128). Amsterdam: Elsevier. Chap. 18.

Beattie, J., & Loomes, G. (1997). The impact of incentives upon risky choice experiments. *Journal of Risk and Uncertainty*, *14*(2), 155–68.

Berg, J., Dickhaut, J., & McCabe, K. (2005). Risk preference instability across institutions: a dilemma. *Proceedings of the National Academy of Sciences USA*, *102*(11), 4209–4214.

Berninghaus, S., Güth, W., Levati, M. V., & Qiu, J. (2010, in press). Satisficing in sales competition: experimental evidence. *International Journal of Game Theory*.

Bhatt, M., & Camerer, C. F. (2005). Self-referential thinking and equilibrium as states of mind in games: fMRI evidence. *Games and Economic Behavior*, *52*(2), 424–459.

Bhattacharya, S., & Pfleiderer, P. (1985). Delegated portfolio management. *Journal of Economic Theory*, *36*(1), 1–25.

Blanco, M., Engelmann, D., & Normann, H.-T. (2007). A Within-subject analysis of other-regarding preferences (Working Paper). Royal Holloway, University of London.

Blanco, M., Engelmann, D., Koch, A. K., & Normann, H.-T. (2008). Belief elicitation in experiments: is there a hedging problem? (Discussion Paper No. 3517). Institute for the Study of Labor (IZA).

Blanco, M., Engelmann, D., Koch, A. K., & Normann, H.-T. (2009). Preferences and beliefs in a sequential social dilemma: a within-subjects analysis (Discussion Paper No. 4624). Institute for the Study of Labor (IZA).

Bolle, F., & Ockenfels, P. (1990). Prisoners' dilemma as a game with incomplete information. *Journal of Economic Psychology*, *11*(1), 69–84.

Bolton, G. E., & Ockenfels, A. (2000). ERC: a theory of equity, reciprocity, and competition. *American Economic Review*, *90*(1), 166–193.

Boschloo, R. D. (1970). Raised conditional level of significance for the 2×2-table when testing the equality of two probabilities. *Statistica Neerlandica*, *24*, 1–35.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*, 1–3.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, *117*(3), 817–869.

Clark, K., & Sefton, M. (2001). The sequential prisoner's dilemma: evidence on reciprocation. *Economic Journal*, *111*(468), 51–68.

Costa-Gomes, M. A., & Weizsäcker, G. (2008). Stated beliefs and play in normal-form games. *Review of Economic Studies*, *75*(3), 729–762.

Croson, R. T. A. (2000). Thinking like a game theorist: factors affecting the frequency of equilibrium play. *Journal of Economic Behavior and Organization*, *41*(3), 299–314.

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, *47*(2), 268–298.

Eichberger, J., Kelsey, D., & Schipper, B. (2008). Granny versus game theorist: ambiguity in experimental games. *Theory and Decision*, *64*(2), 333–362.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics*, *75*(4), 643–669.

Engelmann, D., & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review*, *94*(4), 857–869.

Erev, I., Bornstein, G., & Wallsten, T. S. (1993). The negative effect of probability assessments on decision quality. *Organizational Behavior and Human Decision Processes*, *55*(1), 78–94.

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

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*(3), 817–868.

Fehr, D., Kübler, D., & Danz, D. (2008). Information and beliefs in a repeated normal-form game (SFB 649, Discussion Paper 2008-026). Humboldt University, Berlin.

Fischbacher, U. (2007). Z-tree—Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178.

Friedman, D., & Sunder, S. (2004). Risky curves: from unobservable utility to observable opportunity sets (Mimeo).

Gächter, S., & Renner, E. (2010, in press). The effects of (incentivized) belief elicitation in public good experiments. *Experimental Economics*.

Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, *18*(2), 141–153.

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer, & V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63* (pp. 79–93). Göttingen: Ges. für Wiss. Datenverarbeitung.

Haruvy, E., Lahav, Y., & Noussair, C. N. (2007). Traders' expectations in asset markets: experimental evidence. *American Economic Review*, *97*(5), 1901–1920.

Hertwig, R., & Ortmann, A. (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), 508–515.

Holt, C. A. (1986b). Scoring-rule procedures for eliciting subjective probability and utility functions. In P. Goel, & A. Zellner (Eds.), *Bayesian Inference and Decision Techniques* (pp. 279–290). Amsterdam: Elsevier.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655.

Huck, S., & Weizsäcker, G. (2002). Do players correctly estimate what others do? Evidence of conservatism in beliefs. *Journal of Economic Behavior and Organization*, *47*(1), 71–85.

Isaac, R. M., & James, D. (2000). Just who are you calling risk averse? *Journal of Risk and Uncertainty*, *20*(2), 177–87.

Kadane, J. B., & Winkler, R. L. (1988). Separating probability elicitation from utilities. *Journal of the American Statistical Association*, *83*(402), 357–363.

Karni, E. (1999). Elicitation of subjective probabilities when preferences are state-dependent. *International Economic Review*, *40*(2), 479–486.

Karni, E. (2009). A mechanism for eliciting probabilities. *Econometrica*, *77*(2), 603–606.

Karni, E., & Safra, Z. (1995). The impossibility of experimental elicitation of subjective probabilities. *Theory and Decision*, *38*(3), 313–320.

Manski, C. F. (2004). Measuring expectations. *Econometrica*, *72*(5), 1329–1376.

Nyarko, Y., & Schotter, A. (2002). An experimental study of belief learning using elicited beliefs. *Econometrica*, *70*(3), 971–1005.

Offerman, T., Sonnemans, J., & Schram, A. (1996). Value orientations, expectations and voluntary contributions in public goods. *Economic Journal*, *106*(437), 817–45.

Offerman, T., Sonnemans, J., van de Kuilen, G., & Wakker, P. P. (2009). A truth-serum for non-Bayesians: correcting proper scoring rules for risk attitudes. *Review of Economic Studies*, *76*(4), 1461–1489.

Rey-Biel, P. (2009). Equilibrium play and best response to (stated) beliefs in normal form games. *Games and Economic Behavior*, *65*(2), 572–585.

Rutström, E. E., & Wilcox, N. T. (2009). Stated beliefs versus inferred beliefs: a methodological inquiry and experimental test. *Games and Economic Behavior*, *67*(2), 616–632.

Savage, L. J. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, *66*(336), 783–801.

Schlag, K. H. (2010). New and old results not only for experimental game theory (Mimeo). Universitat Pompeu Fabra.

Schlag, K. H., & van der Weele, J. (2009). Eliciting probabilities, means, medians, variances and covariances without assuming risk neutrality? (Mimeo). Universitat Pompeu Fabra.

Selten, R. (1998). Axiomatic characterization of the quadratic scoring rule. *Experimental Economics*, *1*(1), 43–61.

Wilcox, N. T., & Feltovich, N. (2000). Thinking like a game theorist: comment (Mimeo). University of Houston.

Winkler, R. L., & Murphy, A. H. (1970). Nonlinear utility and the probability score. *Journal of Applied Meteorology*, *9*(1), 143–148.