JENA ECONOMIC RESEARCH PAPERS

2010 – 021

Applying Quadratic Scoring Rule transparently in multiple choice settings: A note

by

Florian Artinger
Filippos Exadaktylos
Hannes Koppel
Lauri Sääksvuori

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Max Planck Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

Applying Quadratic Scoring Rule transparently in multiple choice settings: A note*

Florian Artinger† Filippos Exadaktylos‡ Hannes Koppel§
Lauri Sääksvuori¶

March 23, 2010

Abstract

The quadratic scoring rule (QSR) is often used to guarantee an incentive compatible elicitation of subjective probabilities over events. Experimentalists have regularly not been able to ensure that subjects fully comprehend the consequences of their actions on payoffs given the rules of the games. In this note, we present a procedure that allows the transparent use of the QSR even in multiple-choice scenarios. For that purpose, two methodological means are applied: an alternative representation of the score and a short learning period to familiarize subjects with the payoff mechanism. The results suggest that both means were necessary and successful in facilitating subjects' understanding of the rule.

JEL codes: D84, C90
Keywords: Quadratic scoring rule, experimental methodology, experimental design

*The authors are thankful for helpful comments from participants at the Alhambra Experimental Workshop 2009. Of course, the usual disclaimer applies.

†Max Planck Institute for Human Development and Technical University Berlin, artinger@mpib-berlin.mpg.de

‡University of Granada, exadaktylos@ugr.es

§Max Planck Institute of Economics, koppel@econ.mpg.de

¶Max Planck Institute of Economics, saaksvuori@econ.mpg.de

1 Introduction

The accurate elicitation of subjective expectations has become an important methodological concern in experimental economics. The Quadratic Scoring Rule (henceforth QSR) is currently the most widely used elicitation method.[1] Part of the success of the rule is its theoretical incentive compatibility[2] satisfying the principles set by Induced Value Theory, namely, monotonicity, salience and dominance (Smith 1976). Of particular focus in this paper is the issue of *salience*: that subjects understand the payoff consequences of their actions given the rules of the game. Due to the mathematical formulation underlying the QSR, it is not a trivial matter for the experimental subjects to infer what a certain action means for their payoffs. However, should the subjects not understand the payoff consequences resulting from the formula, the comparative advantage of the mechanism no longer holds. When using the QSR, researchers face a tradeoff between keeping the experiment as simple as possible vs. maintaining salience of the payoff mechanism.

The possible loss of experimental control due to methodological complexity is not a new concern. Other well-known examples include the Becker-DeGroot-Marschak mechanism (Becker et al. 1964) and the elicitation of bids in a Vickrey auction (Vickrey 1961). Albeit theoretically incentive-compatible, experimental subjects often find it difficult to correctly understand the incentives these mechanisms offer.[3] With respect to the QSR, the old concern is illustrated anew in Read (2005, pg. 273): "*I suspect that participants either more-or-less ignore the rule [QSR] or else get so caught up in understanding it that it becomes the focus of their activity*".

In the face of these difficulties, researchers have come up with various responses: Some dropped incentivisation altogether, simply asking subjects to report their beliefs truthfully.[4] However, Gächter and Renner (2006) find that even though the distribution of be-

[1]The QSR was first introduced by Brier (1950), while McKelvey and Page (1990) were the first to apply QSR in an experimental setting in economics. Offerman et al. (1996), in a widely cited article, and Nyarko and Schotter (2002), who showed that elicited beliefs are better than inferred ones, popularized the use of QSR among experimental economists.

[2]Strictly speaking, QSR is an incentive-compatible method only under risk neutrality and the no-stake condition (Kadane and Winkler 1988). Offerman et al. (forthcoming) and Andersen et al. (2009) address the issue of more general risk attitudes and indicate how reported probabilities can be accommodated to include various risk profiles while Palfrey and Wang (2009) experimentally examine whether having a stake in the outcomes affects the subjective probabilities. The analysis provided in this note is a separate issue.

[3]For a more general discussion on cognitive transparency in experimental decision making, see Harrison (1992) and Irwin et al. (1998).

[4]For example, Ivanov (2009) explicitly chose an unconditional lump sum payment over QSR "because it is easy to explain to subjects" (pg.30). Furthermore, the use of lump sum payment is often justified based upon evidence suggesting that incentivisation does not significantly improve stated beliefs, mainly citing Sonnemans and Offerman (2001) (see also Offerman et al. (forthcoming), Dufwenberg and Gneezy

liefs is unaffected, incentives improve the accuracy of the stated beliefs. Other researchers used an elicitation procedure that is simpler to explain to subjects, at the expense of abandoning the properness of the rule (its incentive compatibility).[5] However, properness is not only an important theoretical issue.[6] Palfrey and Wang (2009) demonstrate experimentally that proper linear as well as quadratic scoring rules produce significantly more reliable data than improper ones. Likewise, applications of QSR using frequencies instead of probabilities fail the criterion of properness.[7] When beliefs are elicited in a multiple-choice setting rather than a binary one, applying QSR becomes an even more complicated task.[8]

Even in cases where researchers decide to employ QSR in its proper form, subjects are not encouraged to gain an adequate understanding of the consequences their actions would have on payoffs. The common practice of explaining the rule to subjects is by stating that "it is in their best interest to state beliefs truthfully" and that "it is not important to have a mathematical insight into the formula".[9] However, this deters from the actual incentives the QSR provides and emphasizes that subjects trust the experimentalist's statement. In that way, QSR may be used for payment purposes but does not incentivize subjects'

(2000), Friedman and Massaro (1997), Ortmann et al. (2000), Guarino et al. (2006))

[5]For example, Charness and Dufwenberg (2006) pay 5 Dollar only to subjects whose guesses were close to the realized percentage, whereas Ferraro (2007) rewarded only the best guess with 10 Dollar. Both papers suggest that even if these two elicitation procedures are not incentive compatible - like the more complicated QSR - they still provide adequate incentives for a truthful revelation. In other cases experimental economists were reluctant to drop the QSR. Being aware of the complexity of the rule, however, they employ a simplified version of QSR (see, for example, Croson (1999), Croson (2000), Blanco et al. (2008) and Gächter and Renner (2006))

[6]See Savage (1971) for a theoretical discussion.

[7]Subjects can understand frequencies better than probabilities (see for instance Gigerenzer (1991), Hoffrage and Gigerenzer (2000)). However, as Costa-Gomez and Weizsäcker (2008) mention, this fails to meet the criterion of properness. For papers using frequencies see, for example, Rey-Biel (2009), Neugebauer et al. (2009) and Huck and Weizsäcker (2002).

[8]For example, the task involved 3 possible actions in the designs of Ehrblatt et al. (2007), Terracol and Vaksmann (2009), Fehr et al. (2008), Offerman (2002) and Rey-Biel (2009), 9 possible actions in Leon-Mejia and Miller (2007), and 10 in Schotter and Sopher (2006) and Güth et al. (2009).

[9]For example, Offerman et al. (1996) instruct subjects in the following way: "This payoff is calculated on the basis of a formula. It is not important that you have (mathematical) insight into this formula. But it is important to know that your expected PROBABILITY-PAYOFF is maximized if you report probabilities truthfully. It is to your advantage to report probabilities honestly. It will never be the case that your PROBABILITY-PAYOFF is negative. Other participants will not get to know your reported probabilities, just like you will not know which probabilities are reported by the others. For completeness the formula will be given in a handout. People who want to check the mathematical proof of the statement that for this formula the expected PROBABILITY-PAYOFF is maximized if the probabilities are reported seriously and honestly, can get this proof after the experiment."

actions properly.

In this note, we introduce a novel experimental procedure with the following aims: (i) to ensure the transparency of the QSR while retaining salience and (ii) to generalize its applicability to elicitation tasks with more than two events. This is accomplished by two methodological means: by changing the representation of the mechanism and by introducing a short training period in which subjects make hypothetical choices and receive feedback.

It is well documented in experimental economics that the representation of a task, for instance the stimulus display, the instructions and the framing of the choices, can have a significant effect on subjects' understanding of the decision task (Camerer and Hogarth 1999, for a review). Probably the most well known example is the Wason card task in which subjects' understanding is improved when framed in a real-life context (Gigerenzer and Hug 1992). In a more recent study, Chou et al. (2009) demonstrate the effects that instructions and procedures can have on subjects behavior. They provide evidence on how changing the representation, the instructions and the script help subjects to find the Nash equilibrium. Similarly, experience, learning, and feedback have important effects on subjects' performances, particularly in difficult tasks. Smith and Walker (1993) show the importance of feedback that reveals the link between actions and payoffs in non-transparent situations. Camerer and Hogarth (1999) recapitulate: "it is hard to imagine that incentives alone, without feedback about the quality of previous decisions, would have much effect."

Keeping in mind these well-documented behavioral regularities, we deliver a simple medium to implement the QSR that ensures the comprehension of its reward structure. Furthermore, we provide a representation of the scoring rule that allows researchers to apply the QSR to multiple-choice scenarios. So far, the full payoff schedule has been revealed in binary decision tasks only. We are not aware of any earlier study that explicitly tries to establish the connection between the decision and reward applying the QSR in more than a binary choice task. Our results demonstrate that both means - changing the representation of the mechanism and introducing a short training period - are necessary and successful in facilitating subjects' understanding of the QSR.

This note is organized as follows: In the next section we analyze the QSR and its properties. In section 3, an alternative representation of the QSR in an experiment is delivered; we discuss the design and the procedures of the experiment we run. Results from the experiment are presented in section 4, followed by our conclusions in Section 5.

2 The Quadratic Scoring Rule

A scoring rule is a means to elicit the beliefs of individuals (more formally, these individuals can be called assessors) about uncertain events in the form of subjective probabilities.

This involves computing a score based on the reported assessment that allows the evaluation of the assessors: how accurately they judge the objective probability distribution. The score is derived from the squared distance between the predicted distribution and the observed relative frequency distribution. Formally, assessors report their probability distribution $p = (p_1, \ldots, p_n)$ where $p_i (1 \leq i \leq n)$ refers to the probability that an event i occurs. The incentivisation mechanism that the QSR implements on the following formula:

$$Q_j(p) = \alpha + 2\beta p_j - \beta \sum_{i=1}^{n} (p_i)^2, \tag{1}$$

where j is the event that actually occurs. Starting with a positive threshold, subjects are rewarded for the correct assessment and penalized for every mistake by the squared distance between the reported probability and the observed probability (1 if the event occurs, otherwise 0). Parameters α and β are chosen by the experimenter in order to determine the range of possible payoffs. According to the formula the highest possible outcome is $\alpha + \beta$ and the lowest one $\alpha - \beta$. As can be seen from equation (1), the final score depends not only on the probability assessors place on the event that occurs, but on the entire reported distribution. This property which is not shared by all scoring rules, is crucial when eliciting beliefs in an environment with more than a binary choice.

Consider the equation (1) re-written as:

$$Q_j(p) = (\alpha + \beta) - \beta(1 - p_j)^2 - \beta \sum_{i \neq j} p_i^2. \tag{2}$$

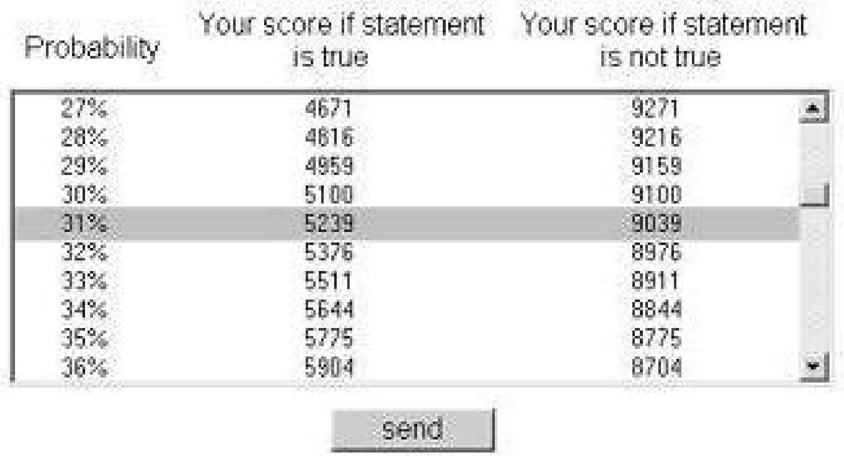
Equation (2) serves to illustrate how the QSR motivates truthful probability assessments. Subjects start with an individual endowment and are penalized for (i) not assigning maximum probability to the event that occurs and (ii) placing a positive probability on events that do not occur. In the case of a binary choice, equation (2) can be further reduced to $(\alpha+\beta)-2\beta(1-p_j)^2$ or equivalently to a prospect $(\alpha+\beta-2\beta(1-P_j^2)_E(\alpha+\beta-2\beta p_j)^2$.[10] This representation allows the rewards generated by QSR to be presented in more readily understandable fashion in a table, as displayed in figure 1.[11]

Subjects are required to state a probability about the uncertain event E while a computer screen similar to figure 1 is displayed to them. The leftmost column refers to a probability that subjects assign to an event whereas the other columns correspond to a score in the case that event E occurs (middle column) or in that E^c occurs (the right-

[10] Notation $x_E y$ depicts a prospect that yields outcome x if event E occurs and outcome y if E^c occurs, with E^c being the complementary event.

[11] Figure 1 is taken directly from Offerman et al. (forthcoming).

most column). Consequently, subjects can directly observe the score for every possible probability assessment.



Probability	Your score if statement is true	Your score if statement is not true
27%	4671	9271
28%	4816	9216
29%	4959	9159
30%	5100	9100
31%	5239	9039
32%	5376	8976
33%	5511	8911
34%	5644	8844
35%	5775	8775
36%	5904	8704

send

Figure 1: The QSR representation as applied in Sonnemans and Offerman (2001)

This representation of payoffs is suitable only in binary choice situations. In more than binary choice tasks this table cannot be produced, for instance in the elicitation of beliefs regarding various political parties standing for election, various possible health states following an operation, various possible end results in a sports event and in many other cases where the action set is not restricted to a binary choice. This is also the case for the majority of the games used in experimental economics (the Dictator Game, the Ultimatum Game, the Public Goods Game, etc.).

We have shown how researchers aim to elicit subjective probabilities via the QSR. Similarly the score is applicable for evaluating the accuracy of beliefs held by a subject. The QSR has a third, traditionally overlooked, potential use. It can serve as an instrument to improve the quality of probability judgments through learning and feedback. In the words of Winkler and Murphy (1968): "*Thus, the results obtained by scoring the assessments, when used as feedback, could serve as a learning device. Specifically, experience should help an assessor to understand the correspondence between judgments and probabilities to confirm the fact that he should set* $\mathbf{r}$ *equal to* $\mathbf{p}$ *in order to maximize his expected score*".[12] This observation creates the conceptual basis for our implementation of the QSR and has motivated us to let subjects engage in a learning period before the real decision task.

3 The alternative representation

We conduct an experiment in which 120 participants report their beliefs regarding the expected behavior in the Dictator Game (DG) and the Ultimatum Game (UG). Both

[12]Here $\mathbf{r}$ and $\mathbf{p}$ are row vectors denoting the reported probability distribution and the true judgement of the assessor, respectively.

games involve two players, player A and player B, between whom an endowment is divided. Player A begins with an initial endowment of 90 Experimental Currency Units (ECU). In the DG, player A dictates the division of the endowment in increments of 10 ECU while player B is passive. In the UG, player A proposes a possible distribution of the endowment, again in increments of 10 ECU. Player B either accepts or rejects.[13] In case of a refusal, both players earn nothing. We elicit the subjects' beliefs in all three decision situations: A´s dictate in the DG, A´s offer in the UG and finally B´s minimum acceptance level in the UG.[14] Hence, every choice set consists of ten possible actions so that the true event $\in \{0, 10, \ldots, 90\}$.

Using the QSR as a payoff mechanism, a table as in figure 1 cannot be used in a more than binary choice task anymore. Instead we opt for subjects doing a limited amount of easily done computations by themselves based on table 1. This is not a payoff table that reports the possible final outcomes. Instead it informs subjects about the consequences of assigning a probability to one possible event and the respective payoff if the event occurs or does not occur.

The leftmost column illustrates the probability a subject assigns to a particular event. The center column represents the value added from a correct judgment that corresponds to the first two terms of the de-composed QSR in equation (2): $(\alpha+\beta)-\beta(1-p_j)^2$. The decreasing trend in payoffs is due to the fact that the QSR penalizes the assessor for not assigning 100 percent probability to an event that occurs. The positive payoff in case of no correct prediction is 10 ECU (due to the initial endowment). The rightmost column displays the costs associated with assigning positive probabilities to events that do not occur (βp_i). Adding all these costs will give the third term in equation (2). That means that the table can still be used for multiple-choice tasks.

An example illustrates how the table works: subject i assigns 70% probability to event A, 20% to event B and 10% to event C. Suppose that event A occurs. The effective reward for assigning 70 percent to the event that occurs is 19.10 ECU, as indicated in the middle column. Similarly the cost of assigning positive probabilities to events B and C that do not occur is 0.40 and 0.10 ECU, respectively, and thus the total cost is 0.40 + 0.10 = 0.50 while the final payoff is 19.10 - 0.50 = 18.60. Using this way of applying QSR, subjects do not simply choose numbers but instead understand the payoff consequences of their

[13] We employ the strategy method (Selten 1967) whereby player B indicates a minimum acceptance threshold that is compared with the received offer.

[14] Deciding on the order of the two tasks (action and elicitation) depends on the aims of the study. The evidence as to whether belief elicitation affects actions is mixed: Erev et al. (1993), Croson (1999), Croson (2000), Nelson (2003) and Lichtenstein et al. (1982) provide evidence of such an effect while Nyarko and Schotter (2002), Costa-Gomez and Weizsäcker (2008) Guerra and Zizzo (2004) and more recently Gächter and Renner (2006) and Armentier and Treich (2009) do not find significant results. Rutström and Wilcox (2009) report evidence that the act of eliciting beliefs affects choices only when elicitation is intrusive.

Stated probability (in percent)	Choice of partner correctly predicted (in ECU)	Costs for giving probabilities to not chosen actions (in ECU)
100	20.00	10.00
95	19.98	9.03
90	19.90	8.10
85	19.78	7.23
80	19.60	6.40
75	19.38	5.63
70	19.10	4.90
65	18.78	4.23
60	18.40	3.60
55	17.98	3.03
50	17.50	2.50
45	16.98	2.03
40	16.40	1.60
35	15.78	1.23
30	15.10	0.90
25	14.38	0.63
20	13.60	0.40
15	12.78	0.23
10	11.90	0.10
5	10.98	0.03
0	10.00	0.00

Table 1: The alternative representation of QSR

actions. Hence, saliency is ensured.

An intuitive representation using the decomposition helps participants to understand the monetary incentives behind the QSR. To ensure comprehension of the reward structure, a short learning period is introduced. The purpose of this learning period is to familiarize assessors with the reward scheme. In the payoff-relevant situations, the participant can focus on the decision itself without spending time and cognitive resources analyzing the characteristics of the reward structure.

The learning period consists of three increasingly difficult questions and guides subjects through the QSR mechanism step by step. For each question, a subject has a maximum of four attempts. If all attempts fail, a visualized solution providing the relevant calculations is given. The first question requires indicating the amount of ECUs earned for a particular correct prediction. This allows subjects to become familiar with the table. In the second question, assessors have to compute the summed cost of several incorrect predictions. In the third question, subjects have to combine for a given situation the reward for the correct prediction with the costs for assigning probabilities to events that did not occur. This last question constitutes a plausible decision situation that participants might face in the incentivized task and serves as a benchmark of subjects having understood the reward structure.[15]

4 Results

The results indicate that subjects initially have problems understanding how the payoff is determined via the QSR. For the first simple question, 17 percent of the subjects gave a wrong answer in their first attempt. For the third question, 25 percent failed in their first attempt. This result highlights the apparent tradeoff that experimentalists face between salience and simplicity when conducting studies utilizing the QSR.

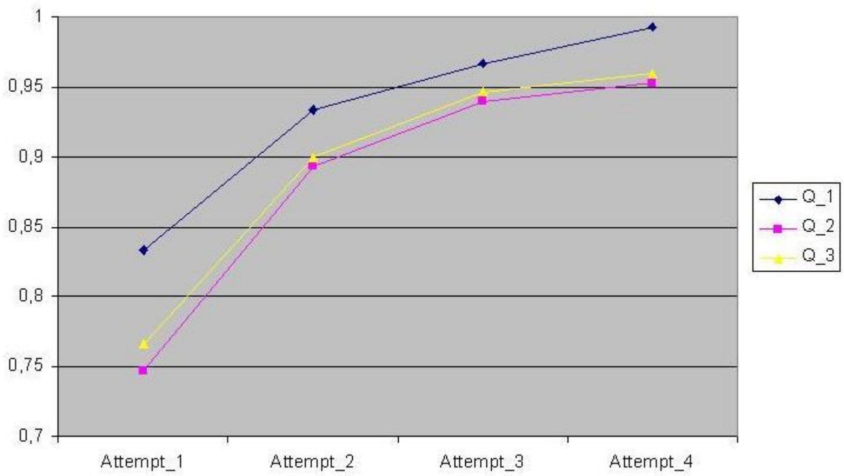


Figure 2: Cumulative share of correct answers

Subjects are very successful in learning how the mechanism works. This requires trial and error before a good understanding is reached. Had the experiment allowed only one attempt to find the correct answer, only 52 percent of all subjects would have been able to find the right answer to all three questions. However, with an average of 1.23 attempts per question, 93 percent of all participants provided the right answer. Most important, 96 percent of experimental subjects gave the right answer for the third question after having had a maximum of four attempts. Figure 2 displays the cumulative share of correct answers across the attempts. In all questions, a significant increase in the

[15]Experimental instructions are available in the appendix of this note. The experimental program in which the learning period is implemented is available from the authors upon request.

amount of correct answers is observed during the learning period. The results illustrate that a proper understanding of the payoff mechanism can be achieved even in a complex elicitation task, providing that participants can familiarize themselves with the reward structure.

5 Conclusions

One of the distinctive features of experimental economics is the commitment to proper financial incentives. This can come at the cost of adding complexity, as in the case of QSR. Experimenters have often neglected to ensure that subjects properly understand the link between their actions and payoffs.

In this note, we provide a means to maintain both salience and ease of understanding. We suggest a modification of how the payoff consequences are represented in an experiment, inducing subjects to discover the properties of the reward scheme themselves. This representation makes it easy to apply the QSR to more-than-binary choice scenarios. The results highlight the importance of practice, learning and feedback in complicated decision tasks. We conclude that incentive compatibility is largely a practical rather than theoretical issue. Simplicity should be a prerequisite.

References

Andersen, S., Fountain, J., Harrison, G. W. and Rutström, E. E.: 2009, Estimating Subjective Probabilities, *Working Paper 09-01*, Department of Economics, College of Business Administration, University of Central Florida.

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

Becker, G., DeGroot, M. and Marschak, J.: 1964, Measuring utility by a single-response sequential method, *Behavioral Science* **9**(3).

Blanco, M., Engelmann, D., Koch, A. and Normann, H.: 2008, Belief elicitation in experiments: is there a hedging problem?, *IZA Discussion Paper No. 3517*.

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

Camerer, C. and Hogarth, R.: 1999, The effects of financial incentives in experiments: A review and capital-labor-production framework, *Journal of Risk and Uncertainty* **19**(1), 7–42.

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

Chou, E., McConnell, M., Nagel, R. and Plott, C.: 2009, The control of game form recognition in experiments: understanding dominant strategy failures in a simple two person guessing game, *Experimental Economics* **12**(2), 159–179.

Costa-Gomez, M. and Weizsäcker, G.: 2008, Stated Beliefs and Play in Normal-Form Games, *Review of Economic Studies* **75**, 729–762.

Croson, R.: 1999, The disjunction effect and reason based choice in games, *Organizational Behavior and Human Decision Processes* **80**, 118–133.

Croson, R.: 2000, Thinking like a Game Theorist: Factors Affecting the Frequency of Equilibrium Play, *Journal of Economic Behavior and Organization* **41**, 299–314.

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

Ehrblatt, W., Hyndman, K., Ozbay, E. and Schotter, A.: 2007, Convergence: An Experimental Study, *mimeo*, New York University.

Erev, I., Bornstein, G. and Wallsten, T. W.: 1993, The negative effect of probability assessment on decision quality, *Organizational Behavior and Human Decision Processes* **55**, 78–94.

Fehr, D., Kübler, D. and Danz, D.: 2008, Information and Beliefs in a Repeated Normal-form Game, *mimeo*, Technical University Berlin, Discussion Paper No. 3627 IZA.

Ferraro, P. C. R.: 2007, Cultural diversity, discrimination, and economic outcomes: an experimental analysis, *Economic Inquiry* **45**, 217–232.

Friedman, D. and Massaro, D.: 1997, Understanding variability in binary and continuous choice, *Psychological Bulletin Review* **5**, 370–389.

Gächter, S. and Renner, E.: 2006, The Effects of (Incentivized) Belief Elicitation in Public-Goods Experiments, *Discussion paper no. 2006-16*, CeDEx.

Gigerenzer, G.: 1991, How to Make Cognitive Illusions Disappear: Beyond Heuristics and Biases, *European Review of Social Psychology* **2**, 83–115.

Gigerenzer, G. and Hug, K.: 1992, Domain-specific reasoning: Social contracts, cheating, and perspective change, *Cognition* **43**(2), 127–171.

Guarino, A., Huck, S. and Jeitschko, T.: 2006, Averting economic collapse and the solipsism bias, *Games and Economic Behavior* **57**, 264–285.

Guerra, G. and Zizzo, D.: 2004, Trust responsiveness and beliefs, *Journal of Economic Behavior and Organization* **55**, 25–30.

Güth, W., Ploner, M. and Regner, T.: 2009, Determinants of in-group bias: Is group affiliation mediated by guilt-aversion?, *Journal of Economic Psychology* **30**(5), 814–827.

Harrison, G.: 1992, Theory and misbehavior of first-price auctions: Reply, *American Economic Review* **82**(5), 1426–1443.

Hoffrage, U., L. S. H. R. and Gigerenzer, G.: 2000, Communicating Statistical Information, *Science* **290**, 2261–2262.

Huck, S. and 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.

Irwin, J., McClelland, G., McKee, M. and Schulze, W.: 1998, Payoff dominance vs. cognitive transparency in decision making, *Economic Inquiry* **36**, 272–285.

Ivanov, A.: 2009, Strategic Play and Risk Aversion in One-Shot Normal-Form Games: An Experimental Study, *mimeo*, Ohio State University.

Kadane, J. and Winkler, R.: 1988, , Separating Probability Elicitation from Utilities, *Journal of the American Statistical Association* **83**, 357–363.

Leon-Mejia, A. and Miller, L.: 2007, The Devil is in the Details ? Sex Differences in Simple Bargaining Games, *Jena economic research papers 2007?069.*

Lichtenstein, S., Fischhoff, B. and Phillips, L.: 1982, Calibration of probabilities: The state of the art to 1980, *in* D. Kahneman, P. Slovic and T. A. (eds), *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, New York, pp. 305–352.

McKelvey, R. and Page, T.: 1990, Public and private information: An experimental study of information pooling, *Econometrica* pp. 1321–1339.

Nelson, W.: 2003, Estimating Subjective Probabilities, *Unpublished Manuscript*, State University of New York at Buffalo.

Neugebauer, T., Perote, J., Schmidt, U. and Loos, M.: 2009, Selfish-biased conditional cooperation: On the decline of contributions in repeated public goods experiments, *Journal of Economic Psychology* **30**, 52–60.

Nyarko, Y. and Schotter, A.: 2002, An Experimental Study of Belief Learning Using Elicited Beliefs, *Econometrica* **70**, 971–1005.

Offerman, T.: 2002, Hurting hurts more than helping helps, *European Economic Review* **46**, 1423–1437.

Offerman, T., Sonnemans, J. and Schram, A.: 1996, Value orientations, expectations and voluntary contributions in public goods, *The Economic Journal* **106**(437), 817–845.

Offerman, T., Sonnemans, J., van de Luilen, G. and Wakker, P.: forthcoming, A truth-serum for non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes, *Review of Economic Studies* .

Ortmann, A., J., F. and Boeing, C.: 2000, Trust, reciprocity, and social history: A re-examination, *Experimental Economics* **3**, 81–100.

Palfrey, T. R. and Wang, S.: 2009, On eliciting beliefs in strategic games, *Journal of Economic Behavior and Organization* **71**, 98–109.

Read, D.: 2005, Monetary incentives, what are they good for?, *Journal of Economic Methodology* **12**(2), 265–276.

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

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

Savage, L.: 1971, Elicitation of Personal Probabilities and Expectations, *American Statistical Association* **66**, 783–801.

Schotter, A. and Sopher, B.: 2006, Trust and trustworthiness in games: an experimental study of intergenerational advice, *Experimental Economics* **9**(2), 123–145.

Selten, R.: 1967, Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments, *Beiträge zur experimentellen Wirtschaftsforschung* **1**, 136–168.

Smith, V.: 1976, Experimental economics: Induced value theory, *American Economic Review* **66**(2), 274–279.

Smith, V. and Walker, J.: 1993, Monetary rewards and decision cost in experimental economics, *Economic Inquiry* **31**(2), 245–261.

Sonnemans, J. and Offerman, T.: 2001, Is the Quadratic Scoring Rule really incentive compatible?, *mimeo*, University of Amsterdam.

Terracol, A. and Vaksmann, J.: 2009, Dumbing down rational players: Learning and teaching in an experimental game, *Journal of Economic Behavior and Organization* **70**, 54–71.

Vickrey, W.: 1961, Counterspeculation, auctions, and competitive sealed tenders, *Journal of finance* **16**, 8–37.

Winkler, R. and Murphy, A.: 1968, Good Probability Assessors, *Journal of Applied Meteorology* **7**(5).

6 Appendix

Instructions

Thank you for coming! You are now about to take part in an experiment on decision making. With taking part in the experiment and reading the following instructions carefully you can earn a considerable amount of money depending both on your own decisions and on the decisions of others.

These instructions and the decisions to be made are solely for your private information. During the experiment you are not allowed to communicate in the laboratory nor with someone outside the laboratory. Please switch off your mobile phone. Any violation of these rules will lead to exclusion from the experiment and all payments. If you have any questions regarding the rules or the course of this experiment, please raise your hand. An experimenter will assist you privately.

The experiment consists of one computerized questionnaire and three separate sections with varying decision tasks. Answering carefully all the items in the questionnaire will earn you four (4) Euros. In each of the three separate sections, one randomly chosen decision determines your earnings from the section. Your overall income from the experiment will be based on the sum of earnings from the three separate sections and the questionnaire. It is in your best interest to make a careful decision in all possible situations. Neither during nor after the experiment will you or any other participant be informed about the true identity of a person with whom you are interacting. Your earnings will be paid privately in cash at the end of the experiment.

During the experiment all decisions and transfers are made in Experimental Currency Units (ECUs). Your total income will be calculated in ECUs and at the end of the experiment converted to Euros at the following rate:

10 ECUs = 1.5 Euro.

The experiment begins with the questionnaire. You have one decision to be taken per computer screen. Please bear in mind that after the introductory stage of two computer screens you have up to 15 seconds to make your decision in each screen. The remaining time is displayed on your screen in the upper right hand corner.

First Section

The first section consists of two decision tasks in which your earnings depend both on your own decisions and one randomly chosen participant. There are two types of individuals: Type A and type B. You will act in both roles. To calculate your earnings from the section only one decision will be randomly chosen. The random decision is determined by the computer at the end of the experiment.

First Decision Task

There are two types of individuals: Type A and type B. Person A decides how to divide a pie of 90 ECUs between him/herself and person B. Person B is passive in this situation. The division is possible in intervals of 10 currency units. Person A can accordingly allocate 0, 10, 20, 30, 40, 50, 60, 70, 80, or 90 ECUs to person B.

Example: Should the person A allocate 30 ECUs to person B, person A earns (90 - 30 =) 60 ECUs.

Second Decision Task

Person A decides how to divide a pie of 90 ECUs between him/herself and person B. Type B person may now either accept or decline the proposed division. Should the person B accept the division, earn both persons ECUs in compliance with the proposed division. Should the person B decline the offered allocation, earn both persons nothing. To determine the final allocation from the second decision task indicates person B the minimum amount of ECUs that he/she is willing to accept. Both the division and the indication of acceptance are possible in intervals of 10 currency units.

You are asked to make your decision in both roles: as a person A and B. Your pay-off relevant decision will be randomly determined by the computer at the end of the experiment.

Example: Person A decides to offer 30 ECUs to person B and thereby keep 60 ECUs for him/herself. Person B indicates the minimum amount of ECUs he/she is willing to accept. Should the amount be smaller or equal to 30 ECUs, receives person B 30 ECUs and person A 60 ECUs. Should the acceptable amount be greater than 30 ECUs, both person receive 0 ECUs from the decision task.

Second Section

The second section consists of three decision tasks that are described below. In this section your earnings depend both on your own decisions and on the decisions of others. At the end of the experiment computer will randomly determine one of the three tasks that will be solely used to assign your earnings from the second section.

There were three situations in which 90 ECUs were at stake:

1. Person A allocates ECUs between him/herself and person B, B is passive;
2. Person A allocates ECUS between him/herself and person B, B is active;
3. Person B indicates the smallest amount that he/she is willing to accept.

All decisions were to be made in intervals of 10 ECUs.

In the following three decision tasks your earnings will be determined by the accuracy of your probability assessment. Your task is to indicate the likelihood that a randomly chosen person has chosen one of the ten possibilities. Please note that the sum of your probability assessments needs to equal 100 per cent.

Your earnings will be calculated on the basis of the following figure 3. A more detailed explanation will follow.

Your probability assesment	Your earnings from a correct prediction	Your cost from an incorrect prediction
100%	20,00 ECU	10,00 ECU
95%	19,98 ECU	9,03 ECU
90%	19,90 ECU	8,10 ECU
85%	19,78 ECU	7,23 ECU
80%	19,60 ECU	6,40 ECU
75%	19,38 ECU	5,63 ECU
70%	19,10 ECU	4,90 ECU
65%	18,78 ECU	4,23 ECU
60%	18,40 ECU	3,60 ECU
55%	17,98 ECU	3,03 ECU

Your probability assesment	Your earnings from a correct prediction	Your cost from an incorrect prediction
50%	17,50 ECU	2,50 ECU
45%	16,98 ECU	2,03 ECU
40%	16,40 ECU	1,60 ECU
35%	15,78 ECU	1,23 ECU
30%	15,10 ECU	0,90 ECU
25%	14,38 ECU	0,63 ECU
20%	13,60 ECU	0,40 ECU
15%	12,78 ECU	0,23 ECU
10%	11,90 ECU	0,10 ECU
5%	10,98 ECU	0,03 ECU
0%	10,00 ECU	0,00 ECU

Figure 3: Earnings table

The payoff consequences of your choice will be explained through an example: Assume a situation in which person A decides how to allocate a pie of 90 ECUs between him/herself and a person B. Person B is passive.

First column in the table contains the probability that you want to assign for a certain possible division. Should you for instance assess that all the 10 possible divisions (from 0 ECUs to 90 ECUs) are equally likely to occur, your decision is to set 10 per cent probability to all possible events.

Second column in the table indicates your earnings from a correct prediction given your probability assessment. Think of the example in which all possible events were assessed to be equally likely and received a probability estimate of 10 per cent. You have inevitably made a correct prediction which earns you 10.90 ECUs.

You have to bear the costs from incorrect probability assessments (third column). In this example you have set 10 per cent probability also for all the events that did not occur.

These incorrect predictions are all associated with a deduction 0.10 ECUs as can be read from the third column in the table.

That is, your total earnings from the task are 10.90 ECU - 0.10 ECU - 0.10 ECU - 0.10 ECU - 0.10 ECU - 0.10 ECU - 0.10 ECU - 0.10 ECU - 0.10 ECU - 0.10 ECU = 11.90 ECU – 9 * 0.10 ECU = 11.00 ECU.

Another example: Assume that you have made following probability assessments: 20% for 0 ECUS, 40% for 10 ECUs, 10% for 20 ECUs and 15% for 30 and 40 ECUs. The randomly chosen person A decides to allocate 10 ECUs to person B. Your probability assessment for that event was 20%. Your earnings from the decision task will be calculated as following: 13.60 ECUs (20% for a correct prediction) - 1.60 ECUs (40% for an incorrect prediction) - 0.10 (10% for an incorrect prediction) - 2 * 0.23 ECUs (two times 15% for an incorrect prediction) = 11.44 ECUs.

Pay attention to the fact that under the given payoff scheme the worst possible monetary outcome happens when you set 100 per cent probability for an event that does not occur. Your earnings in such case would be 0 ECUs. On the contrary, should you set 100 per cent probability for an event that occurs, your earnings would be the highest possible (20 ECUs).

Please note that you are not bound to make your probability assessments in intervals of 5 per cent. This limitation is only for an illustration. That is, you can for instance set a probability of 97% for a certain event. You will receive a complete payoff table once we begin the experiment.

Training Period

Please answer the following control questions. These will give you the opportunity to get familiar with the payoff. At the same time it is ensured that you understood the instructions. Please indicate your answers in the relevant field. To proceed click 'continue'. Once all questions will be answered correctly we will start with the actual experiment.

Please answer following questions with the help of the table 2. You are expecting with a probability of 55% that your partner will decide to give you 30 ECU. Your partner decides to actually give you 30 ECU. How many ECU will you receive for this correct prediction?

After four trials the solution (17.98 ECU) and following table 3 are presented:

Please answer following questions with the help of the above table 2: You are expecting with a probability of 50% that your partner will decide to give you 30 ECU, with a probability of 30% 20 ECU, and with a probability of 20% that he will decide to give you 10 ECU. Your partner decides to actually give you 40 ECU. How high are your costs (in ECU) for this incorrect assessment?

Stated probability (in percent)	Choice of partner correctly predicted (in ECU)	Costs for giving probabilities to not chosen actions (in ECU)
100	20.00	10.00
95	19.98	9.03
90	19.90	8.10
85	19.78	7.23
80	19.60	6.40
75	19.38	5.63
70	19.10	4.90
65	18.78	4.23
60	18.40	3.60
55	17.98	3.03
50	17.50	2.50
45	16.98	2.03
40	16.40	1.60
35	15.78	1.23
30	15.10	0.90
25	14.38	0.63
20	13.60	0.40
15	12.78	0.23
10	11.90	0.10
5	10.98	0.03
0	10.00	0.00

Table 2: The alternative representation of QSR

After four trials the solution $(2.50 + 0.90 + 0.40 = 3.80)$ and following table 4 is presented:

Please answer following questions with the help of the above table 2: You are expecting with a probability of 50% that your partner will decide to give you 30 ECU, with a probability of 30% that he will give you 40 ECU, with a probability of 10% that he will give you 20 ECU, and with a probability of 5% that he will give you 10 ECU respectively 50 ECU . Your partner decides to actually give you 30 ECU. How high is your payoff (in ECU)?

After four trials the solution $(17.50 - 0.90 - 0.10 - 0.03 - 0.03 = 17.50 - 0.90 - 0.10 - 2 * 0.03 = 16.44)$ and following table 5 is presented:

Stated probability (in percent)	Choice of partner correctly predicted (in ECU)	Costs for giving probabilities to not chosen actions (in ECU)
100	20.00	10.00
95	19.98	9.03
90	19.90	8.10
85	19.78	7.23
80	19.60	6.40
75	19.38	5.63
70	19.10	4.90
65	18.78	4.23
60	18.40	3.60
55	**17.98**	3.03
50	17.50	2.50
45	16.98	2.03
40	16.40	1.60
35	15.78	1.23
30	15.10	0.90
25	14.38	0.63
20	13.60	0.40
15	12.78	0.23
10	11.90	0.10
5	10.98	0.03
0	10.00	0.00

Table 3: The alternative representation of QSR

Third Section

In the following decision task your earnings depend only on your own decisions and a random procedure. Your task is to decide between option A and B in ten different situations. At the end of the third section the computer will roll a dice twice (numbers on the dice are between 1 and 10). The first roll determines one of the ten situations and the second roll your earnings from the situation dependent on your choice. In all ten situations there two options available: option A and option B. Both options may earn you a certain amount of ECUs. Look at the situation one - equal to a situation in which the first dice roll turns out to be 1 - displayed in figure 4. Now Option A pays you 20.00 ECUs if the second throw of the ten sided die is 1, and it pays 16.00 ECUs if the throw is 2-10. Option B yields 38.50 ECUs if the throw of the die is 1, and it pays 1 ECU if the

Stated probability (in percent)	Choice of partner correctly predicted (in ECU)	Costs for giving probabilities to not chosen actions (in ECU)
100	20.00	10.00
95	19.98	9.03
90	19.90	8.10
85	19.78	7.23
80	19.60	6.40
75	19.38	5.63
70	19.10	4.90
65	18.78	4.23
60	18.40	3.60
55	17.98	3.03
50	17.50	**2.50**
45	16.98	2.03
40	16.40	1.60
35	15.78	1.23
30	15.10	**0.90**
25	14.38	0.63
20	13.60	**0.40**
15	12.78	0.23
10	11.90	0.10
5	10.98	0.03
0	10.00	0.00

Table 4: The alternative representation of QSR

throw is 2-10. The other situations are similar, except that as you move down the table, the chances of the higher payoff for each option increase.

	Option A: Number of the dice	Option A: Earnings	Option B: Number of the dice	Option B: Earnings	Your decision: Option A or Option B
1.	1	20.00 ECU	1	38.50 ECU	
	2 - 10	16.00 ECU	2 - 10	1.00 ECU	

Figure 4: Situation 1

Example (in case the first throw shows number one):
Assume that the result form the second throw is number one. Should you have chosen option A, your earnings is 20.00 ECUs. Should the second throw be 2, 3, 4, 5, 6, 7, 8, 9

Stated probability (in percent)	Choice of partner correctly predicted (in ECU)	Costs for giving probabilities to not chosen actions (in ECU)
100	20.00	10.00
95	19.98	9.03
90	19.90	8.10
85	19.78	7.23
80	19.60	6.40
75	19.38	5.63
70	19.10	4.90
65	18.78	4.23
60	18.40	3.60
55	17.98	3.03
50	**17.50**	2.50
45	16.98	2.03
40	16.40	1.60
35	15.78	1.23
30	15.10	**0.90**
25	14.38	0.63
20	13.60	0.40
15	12.78	0.23
10	11.90	**0.10**
5	10.98	**0.03**
0	10.00	0.00

Table 5: The alternative representation of QSR

or 10, your earnings are 16 ECUs. Should you have chosen option B, you earnings would be 38.59 ECUs. Should the result form the second throw, however, be 2, 3, 4, 5, 6, 7, 8, 9 or 10 and your decision option B, your earnings would be 1 ECU.

As indicated above, computer will roll a dice twice. The first throw determines one of the ten situations the second throw your earnings from the situation depending on your own decision.

Please answer the following questions concerning some personal details. We will prepare your payment simultaneously. After finishing the questionnaire your final payment will be displayed on your computer screen. You will find out the payoff relevant situations that the computer has chosen in each of three sections.

Thank you for your participation!