A note on the salient implementation of the quadratic scoring rule

Florian Artinger

Max-Planck Institute for Human Development & Warwick Business School, University of Warwick, UK

Filippos Exadaktylos

School of Agriculture Policy and Development, University of Reading, Reading, UK
Department of Economics, Business School, Middlesex University London, London, UK

Hannes Koppel

University of Heidelberg, Department of Economics

Lauri Sääksvuori

National Institute for Health and Welfare, Centre for Health and Social Economics, Helsinki, Finland

---

Abstract

The quadratic scoring rule (QSR) is often used to assess the predictive accuracy of forecast models and guarantee incentive compatible elicitation of subjective probabilities in economic experiments. However, the complexity of the QSR renders it questionable whether non-specialists and experimental participants understand the correspondence between their forecasts and incentives determined by the QSR. This note introduces an explicit implementation procedure to ensure that individuals understand the correspondence between their forecasts and incentives determined by the QSR even in multiple choice scenarios. We show using experimental data that understanding the incentives for truthful reporting does not affect the accuracy of forecasts but may increase the variance of reported distributions.

Keywords: beliefs, experiment, forecasts, incentives, quadratic scoring rule

---

☆Financial support from the Spanish Ministry of Science and Innovation (ECO2013-44879-R) and the Government of Andalusia, Project for Excellence in Research, (P012.SEJ.1436) is gratefully acknowledged.

1. Introduction

Beliefs over uncertain events and other agents' strategies are central building blocks of economics and decision theory. Consequently, the elicitation of subjective probabilities over uncertain events creates an important methodological question germane to the theory and empirics of decision making. The Quadratic Scoring Rule (QSR) is one of the most widely used methods to elicit subjective beliefs over uncertain events in economic experiments. The popularity of the QSR largely derives from its incentive compatibility.[1] However, the mathematical complexity of the QSR renders it questionable whether participants in typical economic experiments understand the correspondence between their actions and payoffs determined by the QSR.

This note introduces an explicit implementation procedure to ensure that individuals understand the correspondence between their forecasts and payoffs determined by the QSR. Moreover, this procedure generalizes the applicability of the QSR to belief elicitation tasks with more than two events. We show using questionnaire data that the vast majority of participants in our experiment achieve to understand the correspondence between their forecasts and payoffs. We show using experimental data that an explicit implementation of the QSR does not affect the accuracy of reported beliefs but may increase the variance of reported belief distributions.

The induced value theory (Smith, 1976) identifies sufficient conditions which allow an experimenter to induce specific characteristics in experimental subjects. A key condition of the induced value theory is salience which is achieved when subjects are directly rewarded based on their choices or performance in an *understandable* fashion. A reward mechanism which is not properly understood by the subjects violates the salience condition and diminishes experimenter's control over subjects' innate characteristics.[2]

The question whether the elicitation of subjective beliefs over uncertain events using the QSR achieves salience in typical economic experiments is primarily empirical. There is, however, very little systematic empirical and methodological research investigating how well participants understand the incentives provided by the QSR. Moreover, there is no empirical research testing whether understanding the incentives for truthful revelation affects the nature of reported beliefs in economic experiments.

Salient implementation of the QSR becomes increasingly difficult in situations where the research question calls for a probabilistic belief elicitation over more than two possible events. Researchers interested in directly eliciting subjective beliefs over multiple possible events often face a trade-off between simplicity and the incentive compatibility of the payoff mechanism.[3]

[1]The incentive compatibility of QSR requires risk neutrality and no-stake condition (Kadane and Winkler, 1988). The relevance of these issues is addressed, for instance, by Andersen et al. (2012), Offerman et al. (2009) and Palfrey and Wang (2009), but is not the focus of this note.

[2]In the context of the QSR, this concern has been previously raised by 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".

[3]The possible loss of experimental control due to methodological complexity is not limited to the implementation of the QSR. Other often mentioned examples include the Becker-DeGroot-Marschak mechanism (Becker

In the face of this trade-off, researchers have come up with various responses. Some researchers have abandoned incentivization altogether and simply request subjects to report their beliefs truthfully without any monetary incentives to do so (e.g., Guarino et al., 2006; Ivanov, 2011; Ortmann et al., 2000; Sonnemans and Offerman, 2001). The lack of monetary incentives for the truthful revelation of beliefs is often justified based upon evidence suggesting that proper monetary incentives do not significantly improve the accuracy of stated beliefs (e.g., Friedman and Massaro, 1997; Offerman et al., 2009; Sonnemans and Offerman, 2001). However, Gächter and Renner (2010) find that monetary incentives may improve the accuracy of stated beliefs in a public goods game experiment even if the distribution of beliefs is largely unaffected. In contrast to the practice of using no monerary incentives at all, some researchers have chosen to rely on incentivized elicitation procedures that are simpler to explain at the expense of abandoning the theoretical incentive compatibility of the mechanism. This category of incentives includes for example a fixed reward if the prediction is correct or close enough to the realized outcome (e.g., Charness and Dufwenberg, 2006; Croson, 2000; Gächter and Renner, 2010).

Palfrey and Wang (2009) demonstrate that incentive compatible logarithmic and quadratic scoring rules produce significantly different forecasts than a linear scoring rule. The theoretical incentive compatibility and evidently more accurately calibrated beliefs, arguably, encourage the use of QSR for belief elicitation in economic experiments. However, researchers willing to implement the QSR have often not encouraged their experimental subjects to gain a thorough understanding of the correspondence between their actions and payoffs. A common practice is to explain the rule to experimental subjects by stating that "it is in their best interest to state beliefs truthfully, while it is not important to have a mathematical insight into the formula" (e.g., Offerman et al., 1996). However, this practice may deter subjects from trying to understand their actual incentives, rendering the belief elicitation procedure less salient. Moreover, in this case the quality of the belief data relies on an assumption that experimental subjects trust experimenter's statement about their best interest.

Our explicit implementation of the QSR is inspired by the Cognitive Load Theory (Sweller, 1988; van Merriënboer and Sweller, 2005) and experimental results showing that the representation of a task, the instructions, and the framing of choices as well as feedback about the quality of prior decisions may affect subjects' understanding of the decision task at hand (Camerer and Hogarth, 1999). The Cognitive Load Theory suggests that complex relations should be constructed from simple ones to reduce the load of short-term working memory. Following this suggestion, we decompose the QSR into two simple components. The first component illustrates the relationship between choices and gains. The second component illustrates the relationship between choices and losses. The combination of these two components fully describes the complex incentive structure of the QSR, allowing researchers to apply the QSR transparently in scenarios with multiple possible events. We show that the vast majority of the experimental

et al., 1964) used to measure agents' willingness to pay and the elicitation of bids in a Vickrey auction (Vickrey, 1961).

subjects achieve to understand the correspondence between their choices and payoffs when the QSR is implemented in an explicit manner.

2. An explicit implementation of the Quadratic Scoring Rule

A scoring rule is a means to incentivize the elicitation of forecasts over uncertain events. The rule computes a score based on the difference between reported forecasts and the observed outcome. Thus, it allows evaluating the accuracy of each assessor's prediction about the uncertain events. The QSR derives its score from the squared distance between the predicted and observed probability distributions. Consider a vector $\mathbf{p} = (p_1, \ldots, p_n)$ containing assessors' reported probabilities over n events, where $0 \leq p_i \leq 1$ denotes the probability that an event i occurs. The score is determined according to the following formula:

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

where α and $\beta \geq 0$, and j is the event that actually occurs. It can be seen that the final score does not only depend on the probability that assessors place on the event that occurs but on the spread of the reported distribution. To illustrate how the QSR motivates truthful revelation of probability assessments and how it can be transparently implemented in settings with multiple potential events, let us separate the effects of p_j and p_i:

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

Equation (2) is a decomposition of the QSR and shows that assessors start with an exogenously given amount ($\alpha + \beta \geq 0$). They are penalized for (i) not assigning maximum probability on event j ($p_j < 1$) and (ii) assigning $p_i > 0$ on events, $i \neq j$, that do not occur. Therefore, assessors' maximize their expected score by reporting a probability distribution vector $\mathbf{p}$ which corresponds to their true judgement $\mathbf{q}$. By contrast, any report $\mathbf{p} \neq \mathbf{q}$ leads to a lower expected score.

In the case of two possible events, equation (2) reduces to $\alpha + \beta - 2\beta(1 - p_j)^2$. This property allows to present all necessary information about the correspondence between assigned probabilities and the generated QSR score in a single table (Table 1), where the left column depicts the probability assigned to an event E. The middle column reports the QSR score when event E occurs and the right column reports the score when the complementary event E^c occurs. Consequently, assessors can directly observe the QSR score for every possible probability assessment. However, this type of representation is only possible for situations with two possible events.

Table 1: A payoff table describing the QSR

Probability	Your score if statement is true	Your score if statement is not true
0%	0	10000
⋮	⋮	⋮
30%	5100	9100
31%	5239	9039
32%	5376	8976
33%	5511	8911
34%	5644	8844
35%	5775	8775
⋮	⋮	⋮
100%	10000	0

Note: Values are calculated with $\alpha = \beta = 5000$.

The assessment of potential economic outcomes and strategic decisions is often not limited to situations with two possible events. For instance, the elicitation of beliefs over the success of political parties in elections in a multiparty system and results of sports events illustrate that the multiplicity of potential events is rather a rule than an exception. This also holds for the majority of experimental games. Given the need for subjective probability elicitation in various decision situations with multiple potential events and the current shortcomings in describing the underlying incentives determined by the QSR, we aim to introduce in this note an explicit implementation of the QSR that addresses these concerns.

Our explicit implementation of the QSR is inspired by the Cognitive Load Theory (see van Merriënboer and Sweller, 2005, for a review). The theory points out that higher levels of task comprehension can be achieved by instructions that minimize the cognitive load of individuals' working memory. Given the complexity of the QSR, a complete explanation of the rule at once may overload the working memory of many experimental participants. One way to decrease the cognitive load is to reduce the number of items that have to be simultaneously processed.[4] Our implementation reduces the number of items by splitting the QSR (equation 2) into two components. The first component describes the rewards for assigning $p_j \geq 0$ on event j that actually occurs, whereas the second depicts the costs associated with assigning $p_i \geq 0$ to events $i \neq j$ that do not occur. The combination of these two components provides a full description of the QSR.

The two components of the QSR can be easily explained through an informative table (Table 2) which is based on equation (2)[5]. In our illustration, we chose to set $\alpha = \beta = 10$. The

[4]For example, Pollock et al. (2002) present to subjects a highly complex task. In one treatment, all of the instructions are presented at once. In another treatment, they divide the instructions into smaller blocks that are explained separately and combined in the end. They found that the first treatment resulted in higher cognitive load and ultimately lower comprehension than the second treatment.

[5]Note that the table is not a traditional payoff table. In other words, it does not directly report all potential

fact that $\alpha = \beta$ prevents potential losses, but also results in a positive payoff for assigning 0 percent on event j. The left column shows the probability assigned by an assessor to a particular event j. The middle column depicts the reward if event j occurs and corresponds to the first component in equation (2), $\alpha + \beta - \beta(1 - p_j)^2$. The right column shows the costs associated with assigning probabilities to an event $i \neq j$ that does not occur. In situations with two possible events i and j, the QSR score is directly derived by subtracting the cost of assigning a positive probability to the event that does not occur, i, from the positive payoff for assigning a positive probability to the event j that does occur. With more than two possible events, costs for assigning probabilities to any event $i \neq j$ are summed up as described by the third term in equation (2), $-\beta \sum_{i \neq j} p_i^2$, and subtracted from the positive payoff. This procedure indicates how the table can be used for prediction tasks with more than two possible events.

Table 2: An explicit representation of the QSR with multiple potential events

Probability that event j occurs (in percent)	Reward for correct prediction that event j occurs (in ECU)	Costs for assigning probabilities to event $i \neq j$ (in ECU)
100	20.00	10.00
⋮	⋮	⋮
75	19.38	5.63
70	19.10	4.90
65	18.78	4.23
⋮	⋮	⋮
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

The following simple example illustrates how the table works. Consider an assessor who is expecting that event A occurs with a probability of 70%, that event B occurs with a probability of 20%, and that event C occurs with a probability of 10%. Now suppose that event A occurs, which results in a positive payoff of 19.10 ECUs. The assessor's expectation that events B and C occur with a positive probability cause a cost of 0.40 ECUs and 0.10 ECUs, respectively. The resulting final payoff for stating the given belief distribution is therefore $19.10 - 0.40 - 0.10 = 18.60$ ECUs.

The Cognitive Load Theory also suggests that cognitive load can be decreased by moving parts stored in a working memory to a long-term memory. We facilitate this process by introducing a learning period which encourages subjects to gradually learn the correspondence

payoffs for any choice and possible event.

between their choices and payoffs determined by the QSR. Moreover, if subjects after a considerable effort fail to provide a correct answer to a question testing their understanding, we provide a detailed demonstration showing how to compute payoffs determined by the QSR.

3. Experimental Design

We conduct a survey to test subjects' understanding of the quadratic scoring rule when the rule is described using our explicit implementation procedure. Moreover, we design an experiment to compare reported belief distributions in two different treatments. Our first treatment (TRUST) implements the QSR using a common practice to state that it is in subjects' best interest to truthfully report their beliefs over potential events. Our second treatment (EXPLICIT) implements an explicit representation of the QSR with a learning period before the belief elicitation.

When evaluating the performance of different belief elicitation mechanisms, researchers often face a challenge that the true probability is typically unobservable, yet necessary for the evaluation of the belief elicitation mechanism (see Schlag et al., 2015, for a discussion). Consequently, it has been proposed that inducing beliefs about the objective probabilities of events may offer a practical solution to overcome this problem in evaluation. However, at the same time, research questions often call for the elicitation of subjective beliefs over uncertain strategies of other agents, in which case it is not possible to induce objective beliefs. To test the performance of our explicit implementation of the QSR in these two different domains of belief elicitation, we design an experiment with two different belief elicitation tasks.

3.1. Belief elicitation tasks

In the first task, participants are requested to report their beliefs about the outcome of a draw from an "ambiguous urn" containing a fixed, but unknown proportions of balls A, B, C, D and E. Beliefs about the objective probabilities are induced by letting participants to draw a ball 50 times from the urn with replacement.[6] Furthermore, note that the order and outcome of each draw was the same for each participant. Participants were not allowed to take any notes about the drawn balls. Afterwards participants were requested to report their probabilistic beliefs (in integers) about the outcome of a subsequent draw. In other words, they were requested to forecast whether the next ball drawn from the urn is either ball A, B, C, D, or E. We thereby elicit a full probability distribution of beliefs in a non-strategic situation with an objective probability benchmark.

The second task elicits participants' beliefs about behavior in a previously conducted public goods game experiment. This task addresses belief elicitation in a situation with strategic uncertainty. Our participants did not take part in the public goods game experiment, but

[6]This is in contrast to the simplest way of inducing beliefs by informing participants directly about the objective probability of an event, as for example in Hao and Houser (2012). A practical problem with this approach is that the task of reporting probabilistic beliefs becomes trivial if participants are explicitly informed about the probabilities of different events.

were requested to forecast the decision of a randomly drawn experimental subject participating in the original public goods game experiment. In our experiment, participants received the instructions of the public goods game experiment and had to familiarize themselves with the rules of the experiment. In the original experiment, subjects had to choose an integer between 0 and 10 which they contributed to the public account. This results in 11 possible contribution decisions. We elicit our participants' beliefs about the contribution of a randomly drawn experimental subject in the original experiment by requesting participants to report their beliefs (in integers) over each possible contribution level.

3.2. Treatments

Our treatments employ different procedures to implement the QSR.[7] The treatment TRUST implements a practice where payoffs are calculated based on the QSR, while participants are only told that it is in their best interest to truthfully report their beliefs. Like many other experiments, we offer participants an opportunity to inspect the functional form of the QSR and a mathematical proof showing that truthful reporting maximizes their expected payoff.

In treatment EXPLICIT, we implement the previously described explicit representation of the QSR. This treatment includes also a training period. The exact calculation of the payoff is explained using Table 2 and two simple examples. We implement a training period to familiarize participants with the mechanism translating their reported beliefs into payoffs. The training period consists of three questions. Subjects have a maximum of four attempts for each question. If a subject fails to provide the correct answer in the last attempt, the computer presents a visualized solution providing the relevant calculations. The first question asks for the size of the reward, if the subject places a positive probability on an event that actually occurs. The second question requests to indicate the total cost for assigning positive probabilities to several events that do not occur. Finally, the third question asks for the combination of rewards and costs.

3.3. Experimental Protocol

The experiment was conducted in the experimental laboratory of the Max Planck Institute for Economics in Jena, Germany, using z-Tree (Fischbacher, 2007) and ORSEE software (Greiner, 2015). Participants were mainly undergraduate students from a wide range of academic disciplines. There were 32 participants in treatment TRUST and 31 participants in treatment EXPLICIT.

After entering the computer laboratory participants received written instructions. After this initial phase, participants went through the training period in treatment EXPLICIT. There was

[7] We implement two other treatments in addition to the treatments reported in this note. These treatments implement the treatment EXPLICIT without the training period and using natural frequencies instead of probabilities. The results from these two additional treatments are qualitatively and quantitatively very similar to the results of treatment EXPLICIT and omitted from this note for brevity. Instructions for both treatments reported in this note are made available in the Electronic Supplementary Material. This material also contains an exact wording of the questions presented to the subjects during the training period in treatment EXPLICIT.

no training period in treatment TRUST where participants reported their beliefs immediately after reading the instructions. In addition to the belief distributions, we elicited participants risk attitudes using an incentivized measure for risk aversion developed by Holt and Laury (2002). Participants received feedback concerning the outcomes of the three tasks only after completing the entire experiment. The experiment ended with a short post-experimental questionnaire collecting information about participants' statistical familiarity, cognitive abilities and demographics. At the end of the experiment, participants received their payment based on one randomly chosen belief elicitation task and the risk elicitation task. The experiment lasted about 60 minutes. Participants earned on average 10.62€ with a minimum of 5€ and a maximum of 15€. Participants earnings include a 2.5€ show-up fee.

3.4. Ethics Statement

The study design was reviewed and approved by the departmental institutional review board of the Max Planck Institute for Economics before the study began. The treatment of participants was in agreement with the ethical guidelines of the German Psychological Society. Specifically, all participants gave their written informed consent to participate voluntarily, assuring them that analyses and publication of experimental data would be without an association to their real identities. Moreover, random assignment to visually separated cubicles and private payment at the end of the experiment preserved the anonymity of participants. The experiment involved no deception of participants. As in other socioeconomic experiments, there were no additional ethical concerns.

4. Results

This section summarizes our main empirical findings. First, we evaluate whether participants understand the correspondence between their forecasts and the payoff determined by the QSR. Second, we investigate whether implementing an explicit representation of the QSR changes reported belief distributions.

Result 1: An explicit and gradual description of the QSR leads to a salient implementation of the QSR.

In treatment EXPLICIT, experimental participants were requested to answer questions about the correspondence between certain hypothetical forecasts and payoffs. Participants' answers to these questions serve as a proxy for understanding the incentives provided by the QSR. Figure 1 shows the cumulative distribution of correct answers to questions testing participants' understanding of the incentives provided by the QSR. Figure 1 shows that participants had initial difficulties to understand the correspondence between the presented hypothetical forecasts and payoffs determined by the QSR. Seven participants out of 31 participants gave a wrong answer to the first and simplest question in their first attempt. However, the number of incorrect answers diminishes over time. This suggests that participants effectively learn to

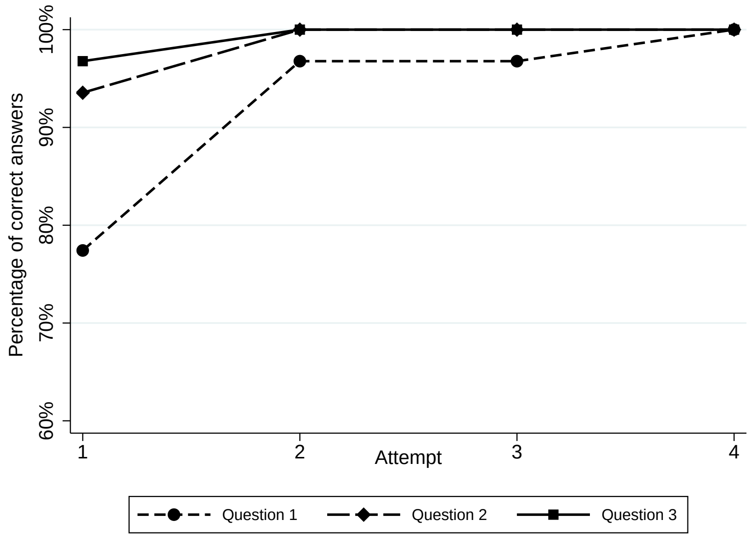

Figure 1: Cumulative distribution of correct answers in treatment EXPLICIT

understand the incentives provided by the QSR through trial and error. Notably, not a single participant gave a wrong answer in the last attempt of any question. However, without having a chance for multiple attempts, only 68 percent of all participants would have correctly answered all three questions. These findings suggest that a proper understanding of the payoff mechanism can be achieved even in complex belief elicitation tasks, if the reward structure is carefully described to participants and they have an opportunity to familiarize themselves with the reward structure before making their actual decisions.

Table 3: Average QSR scores

	Urn		Public Goods		N
	Average	Std.dev	Average	Std.dev	
TRUST	1.198	0.044	0.994	0.150	32
EXPLICIT	1.183	0.046	0.996	0.100	31

Table 4: The distribution of reported beliefs in the ambiguous urn task

Event	True percentage	TRUST		EXPLICIT	
		Average	Std.dev	Average	Std.dev
A	10	11.81	8.36	9.32	8.80
B	20	22.47	8.67	17.06	10.76
C	30	24.81	9.43	33.35	14.21
D	30	28.19	9.96	31.13	11.30
E	10	12.72	7.48	9.13	6.53

To investigate the accuracy of reported belief distributions, we compute an individual QSR score for each participant. This score is based on the reported belief distribution and the actual distribution of events. In other words, we match participants' forecasts with every possible event and compute the average QSR score over all possible events. This average QSR score measures the overall accuracy of reported beliefs for each participant in each task separately.

Result 2: There is no difference in the accuracy of beliefs between treatments TRUST and EXPLICIT in the ambiguous urn task.

As no participant in treatment TRUST showed interest to inspect the functional form of the scoring rule or the mathematical proof showing that truthful revelation maximizes participants' expected payoff, we attribute any potential differences in reported beliefs distributions to differences between the implementation of the QSR. We find that the average QSR score is higher in treatment TRUST than in treatment EXPLICIT in the ambiguous urn task (Table 3). This difference is not statistically significant (Mann-Whitney test: $p = .07$, two-sided; Kolmogorov-Smirnov test: $p = .215$, two-sided). However, given the relatively small sample size, the statistical power of the comparison, $(1 - \beta) = 0.368$, is low.

Result 3: The variance of reported belief distribution is significantly higher in treatment EXPLICIT than in treatment TRUST in the ambiguous urn task.

Table 4 shows that reported beliefs are on average closer to the true objective probability in treatment EXPLICIT than in treatment TRUST for all events except B. This observation indicates that the higher average QSR score in treatment TRUST is not driven by smaller deviations from the true probability distribution in treatment TRUST but by the higher variance of reported belief distributions in treatment EXPLICIT. We test the statistical significance of this observation by computing the variance of reported belief distributions separately for each individual and find that the variance of reported beliefs are significantly higher in treatment EXPLICIT than in treatment TRUST (Mann-Whithey test: $p = .02$, two-sided; Kolmogorov-Smirnov test: $p = .001$, two-sided).

It is noteworthy that a uniform distribution which assigns equal probability to all potential events in the ambiguous urn task yields a QSR score of 1.20. Figure 2 shows the distribution of QSR scores in the ambiguous urn task in both treatments. We observe that half of the participants in treatment EXPLICIT achieve a lower QSR score than what they would have achieved by reporting a uniform distribution. Moreover, we find that there are substantially more participants who achieve a QSR score below 1.20 in treatment EXPLICIT than in treatment TRUST (Fisher's exact test: $p = .081$). Taken together, we observe that there is more heterogeneity in reported belief distributions in treatment EXPLICIT which explicitly describes the incentives for truthful revelation. We are not capable to explain the reasons for this difference in reported beliefs. However, one may conjecture that understanding the incentives for truthful reporting makes participants to think more thoroughly and systematically about the decision task which in turn leads to higher variance of reported belief distributions.

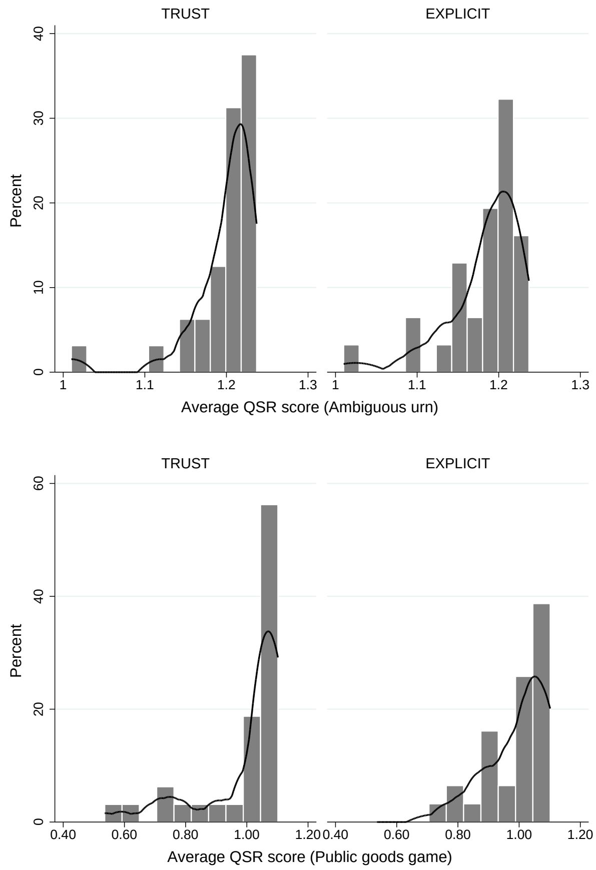

Figure 2: The distribution of QSR scores in the ambiguous urn and public goods game tasks in both treatments. Solid lines show kernel densities of QSR scores using the Epanechnikov kernel function. Bandwidths are calculated as to minimize the mean integrated squared error for an underlying Gaussian density.

Result 4: There is no difference in the accuracy of beliefs or variance of reported belief distributions between treatments in the public goods game task.

We find that there is no significant difference in the average QSR scores between treatments TRUST and EXPLICIT in the public goods game task (Mann-Whitney test: $p = .32$, two-sided: Kolmogorov-Smirnov test: $p = .135$, two-sided). Table 5 shows the reported belief distributions in treatments TRUST and EXPLICIT. We observe that there is hardly any differences in the spread of reported belief distributions between the treatments. We find that the variances of reported belief distributions are not significantly different between treatments TRUST and EXPLICIT in the public goods game task (Mann-Whithey test: $p = .39$, two-sided; Kolmogorov-Smirnov test: $p = .137$, two-sided).

Figure 2 shows the distribution of QSR scores in the public goods game task in both treatments. In the public goods game task, a uniform distribution which assigns equal probability to all potential events yields a QSR score of 1.09. We observe that a vast majority of the participants in both treatments achieve a lower QSR score than what they would have achieved by reporting a uniform distribution. We find that there is no significant difference in the proportion of participants achieving a QSR score below 1.09 between treatments EXPLICIT and TRUST

Table 5: Reported beliefs in the public goods game task

Event	True percentage	TRUST		EXPLICIT	
		Average	Std.dev.	Average	Std.dev.
0	17	18.69	21.59	14.71	16.94
10	2	9.28	6.95	7.19	7.07
20	18	10.88	9.11	7.61	7.04
30	14	12.28	8.28	8.45	7.65
40	9	13.00	9.71	12.26	7.94
50	16	13.13	9.69	19.74	16.17
60	9	6.78	5.59	8.48	7.33
70	0	3.59	3.23	7.81	9.82
80	5	2.69	2.61	3.90	4.41
90	1	2.22	2.42	3.16	5.30
100	9	7.47	18.53	6.68	12.31

(Fisher's exact test: $p = .528$). Taken together, we conclude that there are no differences in the accuracy of beliefs or variance of reported belief distributions between treatments TRUST and EXPLICIT in the public goods game task.

5. Conclusions

One of the distinctive features of experimental economics is the commitment to proper financial incentives. However, this practice can often come at the cost of adding complexity, as in the case of the QSR. We introduce in this note an explicit implementation procedure of the QSR. This procedure maintains the incentive compatibility of the payoff mechanism and facilitates individuals' understanding of the incentives determined by the QSR. Our explicit implementation of the QSR distinguishes between gains from a correct prediction and costs from incorrect predictions. This procedure generalizes the applicability of the QSR to belief elicitation tasks with multiple potential events. Our survey data show that an explicit and gradual description of the QSR facilitates individuals' understanding of the incentives determined by the QSR. We show using experimental data that understanding the incentives does not change the accuracy of reported beliefs but may increase the variance of reported belief distributions in non-strategic tasks.

Our results highlight that individuals' may gain an understanding of complex mathematical relationships during an experimental session if the relation between a set of inputs and outputs is carefully explained. At the same time, our results suggest that a more salient implementation of the experimental task does not necessarily affect individuals' actions.

References

Andersen, S., Fountain, J., Harrison, G. W., Hole, A. R., and Rutström, E. E. (2012). Inferring beliefs as subjectively imprecise probabilities. *Theory and decision*, 73(1):161–184.

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

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

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

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

Gächter, S. and Renner, E. (2010). The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics*, 13(3):364–377.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1:114–125.

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

Hao, L. and Houser, D. (2012). Belief elicitation in the presence of naÃ¯ve respondents: An experimental study. *Journal of Risk and Uncertainty*, 44:161–180.

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

Ivanov, A. (2011). Attitudes to ambiguity in one-shot normal-form games: An experimental study. *Games and Economic Behavior*, 71(2):366–394.

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

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 Kuilen, G., and Wakker, P. P. (2009). A truth serum for non-bayesians: Correcting proper scoring rules for risk attitudes. *The Review of Economic Studies*, 76(4):1461–1489.

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.

Pollock, E., Chandler, P., and Sweller, J. (2002). Assimilating complex information. *Learning and instruction*, 12(1):61–86.

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

Schlag, K., Tremewan, J., and van der Weele, J. (2015). A penny for your thoughts: A survey of methods for eliciting beliefs. *Experimental Economics*, 18:475–490.

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

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

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12:257–285.

van Merriënboer, J. J. and Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational psychology review*, 17(2):147–177.

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