WORKING PAPERS

Simple Belief Elicitation

Karl H. Schlag
James Tremewan

September 2012

Working Paper No: 1204

DEPARTMENT OF ECONOMICS

UNIVERSITY OF VIENNA

Noname manuscript No.
(will be inserted by the editor)

Simple Belief Elicitation

Karl H. Schlag,James Tremewan

Received: date / Accepted: date

Abstract We present methods of belief elicitation which are applicable for any non-trivial utility function. Unlike existing techniques that account for deviations from risk-neutrality, these methods are highly transparent to subjects. Rather than identifying beliefs exactly we identify bounds on beliefs, thus trading off precision for generality and simplicity.

Keywords Belief elicitation

1 Introduction

As with other decision-making tasks in experimental economics, it is deemed desirable to incentivize belief elicitation; that is to pay subjects in such a way that revealing their "true belief" is utility maximizing. The most commonly used incentive mechanism is the quadratic scoring rule (QSR) (Brier, 1950). The standard way of implementing the QSR in a situation with n possible outcomes is to ask subjects to report for each outcome i the probability q_i with which they believe that it will occur. They are then paid according to the following function of the reported probabilities when outcome j has occurred:

$$Q(q,j) = \alpha - \beta (1 - q_j)^2 - \beta \sum_{i \in \{1,..,n\} \setminus \{j\}} q_i^2 \ ,$$

where $\beta > 0$.

There are two major drawbacks of the QSR and any other scoring rule where payment is deterministic and based on a single realized outcome. First of all, the scoring rule cannot be incentive compatible for all utility functions

J. Tremewan
University of Vienna
Tel.: +43-1-4277-37456
Fax: +43-1-4277-9374
E-mail: james.tremewan@univie.ac.at

(Schlag and van der Weele, 2012). Secondly, the implementation can be confusing for subjects not familiar with mathematics and specifically probabilities.

In the first elicitation method, we ask the subject to guess the empirical frequency of a each outcome and we then award a prize if and only if their guess coincides with the realized frequencies. This method has been used before (Wilcox and Feltovich, 2000; Bhatt and Camerer, 2005), however the properties of this method do not appear to have been well understood. References in the literature state only that the modal frequency of outcomes is elicited (Wilcox and Feltovich, 2000; Blanco et al, 2010) or that it is valid only when the true subjective probability coincides exactly with one of the possible empirical distributions (Costa-Gomes and Weizsacker, 2008).

Not only does the method elicit beliefs about the modal frequencies, but we show in addition that it enables the researcher to identify a region on a simplex in which the belief of the subject must lie. Inference does not require postulating any assumptions on the utility function beyond assuming that the subject strictly prefers the prize. For binary events this region is an interval of width $1/(n+1)$, where n is the number of realizations of the variable in question. For sizes of n feasible in laboratory studies this level of precision should be adequate for many practical purposes, and indeed the degree of precision to which people are capable of expressing subjective probabilities is questionable. When asked for percentages, respondents tend to answer in multiples of five [1] (Manski, 2004), which means that there is no loss of precision for a binary event when $n \geq 20$. We show that our rule is most precise in a well defined sense.

The second method discussed in this paper elicits beliefs about the median of a distribution. Subjects are asked to report a number such that half the realizations will be below that number, and are rewarded the prize if and only if this condition holds. As with the first method, this is extremely straightforward to explain to subjects, and is equally valid for all non-trivial utility functions. Similar techniques work for other quantiles.

The applicability of the QSR to only risk neutral subjects has been treated in two different ways: estimating individual utility functions by presenting the subjects with a large number of choices between binary lotteries and using this estimation to adjust the stated beliefs (Offerman et al, 2009) ; using a randomized payment rule, that is paying with binary lotteries (Schlag and van der Weele, 2012). However both of these avenues add to the confusion of the subject which is what we wish to minimize.

Reducing the complexity of instructions, and simplifying the communication of probabilistic information has not been a focus of the experimental economics literature on belief elicitation. Confusion and difficulties with processing probabilities without doubt increase noise and possibly introduce biases in responses. In light of this we suggest that an important route to improving the quality of belief elicitation is by better facilitating the understanding and communication of probabilities by subjects.

[1] Responses do become more precise when the percentages are close to zero or 100.

Probabilities can be expressed in a number of different ways: as a number, a percentage, or as a frequency. There is substantial evidence that even highly educated individuals often perceive mathematically equivalent probabilities as different when presented in the alternative formats. Lipkus et al (2001) found that in a sample where 90 % of respondents had at least some tertiary education, 40% were unable to convert a percentage to a frequency, while 79% were unable to convert a frequency to a percentage. Similar but more extreme results have been found for less educated respondents (Schwarz et al, 1997). Consequently, the format of probabilities has the potential to affect responses when eliciting beliefs.

There is evidence that people tend to be more comfortable and better able to process probabilities expressed as natural frequencies rather than other formats.[2] Experiments by Kahneman and Tversky (1983) find that expressing probabilities as natural frequencies can mitigate the conjunction fallacy, while Gigerenzer and Hoffrage (1995) show that it also facilitates Bayesian reasoning. Cosmides and Tooby (1996) confirm the latter result and argue that human cognitive architecture has evolved to process natural frequencies rather than single-event probabilities in many situations. Schapira et al (2001) report that participants in their study identify frequency formats as being intuitive and easy to interpret. To illustrate the primacy of natural frequency in probability related cognition the reader may try to explain the meaning of the statement "a fair coin will come up heads with probability 0.5" to someone not fluent in mathematics *without referring to natural frequencies*!

Most of the literature on belief elicitation focuses on payments based on single events. However, in many laboratory experiments, there will be not just one but many independent realisations of the random variable of interest. Take, for example, a one-shot prisoners' dilemma experiment where the experimenter is interested in beliefs the subjects hold about the probability of defection. If there are 21 subjects per session, each stated belief can be matched with the 20 realizations of the decisions of others. This allows the experimenter and subject to communicate purely in terms of natural frequencies: "How many of the other 20 participants will choose to defect." The advantage of this approach in increasing transparency for the subjects has been recognized by some experimenters. Blanco et al (2010) ask how many out of nine subjects will cooperate as a second mover in a sequential prisoners' dilemma, but incentivize the guess with the QSR leaving the reported beliefs vulnerable to distortion because of deviations from risk-neutrality.

[2] Another avenue we believe worth pursuing is the use of graphical aids. There has been a great deal of work on this in the fields of cognitive psychology and medical risk communication which could both complement and be complemented by experimental economics methodology. See, for example, references in Schapira et al (2001).

2 Eliciting Probabilities and Functions Thereof

In this section we present a non-randomized method for eliciting probabilities, derive tight bounds on the "true" underlying probabilities and show how this can be used to gain an understanding of perceived means, variation and expected utility.

2.1 Eliciting Probabilities

Let Y be a random variable with k possible outcomes $s_1,..,s_k$, where p_i is a subject's subjective belief about the probability that outcome s_i will occur. Subjects are asked to report $b=(b_1,..,b_k)$, b_i being a non-negative number for all i, and are paid a prize of value R if and only if for all $i\in\{1,...,k\}$ b_i is equal to the number of times s_i occurs out of n independent[3] realisations of Y. We call this the frequency guessing method. In the context of a strategic form game in a laboratory experiment, each s_i is a strategy available to the subject's partner. The subject is then told that they will be awarded a prize if they can correctly guess the number of people in their partner's role who play each strategy. Thus, from the standpoint of the subject making the report, the prize will be awarded with probability

$$f(b)=\frac{n!}{b_1!\cdot..\cdot b_k!}\prod p_i^{b_i}.$$

It follows immediately that the subject maximizes expected utility if and only if they maximize the probability f of receiving the prize. Hence, and without loss of generality, we are interested in the relationship between the maximizers of f and the underlying subjective beliefs. In the following we provide a complete characterization of this relationship.

Let B be the set of feasible reports, so $B=\left\{b\in\{0,1,..,n\}^n : b_i\geq 0\forall i, \sum_{i=1}^n b_i=n\right\}$.

Proposition 1 *Consider $b\in B$. Then b maximises f over all B if and only if*

$$\begin{gathered}\frac{b_i}{b_j+1}\leq\frac{p_i}{p_j}\leq\frac{b_i+1}{b_j}\ \forall j\neq i \text{ when } p_j,b_j\neq 0 \\ b_j=0 \text{ if } p_j=0.\end{gathered}\tag{1}$$

In particular, if b maximizes f then

$$\frac{b_i}{n+k-1}\leq p_i\leq\frac{b_i+1}{n+1} \text{ holds for all } i.\tag{2}$$

[3] The independence assumption will be discussed in Section 4.

Proof To prove the "only if" statement suppose b maximises $f(b)$. If $p_v = 0$ then clearly best if $b_v = 0$. For any $u \neq v$ with $b_v, p_v > 0$,

$$\begin{aligned} & f(b_1, ..., b_u, ..., b_v, ...b_k) - f(b_1, ..., b_u + 1, ..., b_v - 1, ...b_k) \\ = & \frac{n!}{b_1! \cdot .. \cdot b_k!} \prod p_i^{b_i} - \frac{b_v p_u}{(b_u + 1) p_v} \frac{n!}{b_1! \cdot .. \cdot b_k!} \prod p_i^{b_i} \\ = & f(b) \left(1 - \frac{b_v p_u}{(b_u + 1) p_v}\right) \end{aligned}$$

which gives us the set of constraints

$$b_v p_u \leq (b_u + 1) p_v \forall u \neq v. \tag{3}$$

Now $p_i = \sum_j \frac{b_j}{n} p_i = \frac{b_i}{n} p_i + \sum_{j \neq i} \frac{b_j}{n} p_i \leq \frac{b_i}{n} p_i + \sum_{j \neq i} \frac{b_i + 1}{n} p_j = \frac{b_i}{n} + \frac{1}{n}(1 - p_i)$ which implies

$$p_i \leq \frac{b_i + 1}{n + 1}. \tag{4}$$

Also, for $b_i > 0$, $p_i = 1 - \sum_{i \neq j} p_j \geq 1 - \sum_{j \neq i} \frac{(b_j + 1) p_i}{b_i} = 1 - \frac{p_i}{b_i}(n - b_i + k - 1)$, which implies

$$p_i \geq \frac{b_i}{n + k - 1}. \tag{5}$$

To prove the "if" statement assume that b satisfies (1). Consider any b' such that $f(b') > 0$, $b'_u > b_u$ and $b'_v < b_v$. Hence, $p_v > 0$. From the above equations above we obtain

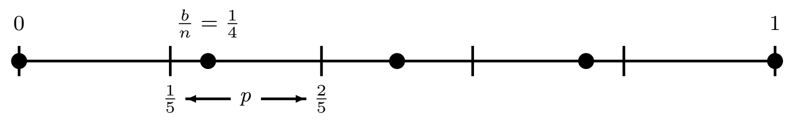
$$\begin{aligned} & f(b'_1, ..., b'_u, ..., b'_v, ...b'_k) - f(b'_1, ..., b'_u + 1, ..., b'_v - 1, ...b'_k) \\ = & f(b') \left(1 - \frac{b'_v p_u}{(b'_u + 1) p_v}\right) \\ > & f(b') \left(1 - \frac{b_v p_u}{(b_u + 1) p_v}\right) > 0. \end{aligned}$$

This means that whenever we increase the report of event u by one and at the same time decrease the report of v by one then the probability of winning the prize goes down, provided the report of u was above b_u and the report of v was below b_v. Thus, for any given p we can compare $f(b)$ to any other $f(b')$, by repeating the above for all $u \in \{i : b'_i > b_i\}$ and $v \in \{i : b'_i < b_i\}$. This shows that b maximizes f over all $b' \in B$ which completes the proof.

Figure 1 demonstrates this result for $k = 2$ and $n = 4$. The dots show the possible reports (divided by n) and the surrounding intervals show the possible values of p given the reports. In the figure we see that only those beliefs on the boundary between two regions give rise to two different optimal reports. More generally, our proof of Proposition 1 reveals that any subject with beliefs that satisfy (1) with strict inequalities has a unique best report.

For $k = 2$ we hasten to point out that one cannot extract more precise information for any given utility function in the following sense. Consider any

Fig. 1 Reported and consistent true beliefs for $k = 2$ and $n = 4$



alternative payment rule with the same input, that is a subject's stated belief about the number of times that an outcome will occur. For a given utility function u let P_b^u be the set of beliefs under which it is optimal under the alternative rule to report b, $b \in \{0, 1, .., n\}$. Then $\cup_{b \in \{0,1,..,n\}} P_b^u = [0, 1]$. Consequently, $\max_{b \in \{0,1,..,n\}} d(P_b^u) \geq 1/(n+1)$ where $d(P_b^u)$ is maximal distance between any two points belonging to P_b^u (where d is its width if P_b^u is an interval). Let the *minimal precision* of a rule be the negative of the maximal difference between any two probabilities that lead to the same report. Then we find that there is no payment rule with a strictly higher minimal precision than the one we have presented. In fact, it is easy to see that the inferred true probabilities of any rule with this value of minimal precision are unique. We summarize.

Proposition 2 *Any alternative rule that elicits the frequency of the occurrence of a single event (so $k = 2$) has a strictly lower minimal precision than that of the frequency guessing method.*

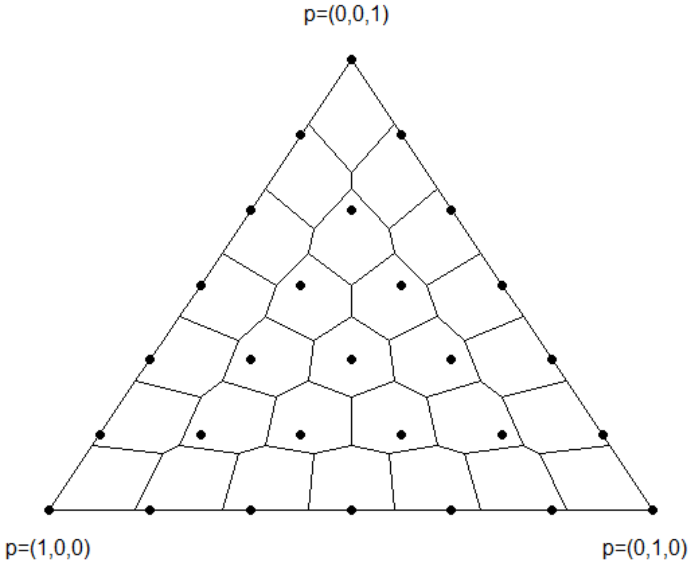
In general the set of feasible probabilities is constrained by $p_i \geq 0$ for all i, by $\sum_{i=1}^{n} p_i = 1$ and by the constraints given in (1). Figure 2 shows how these constraints divide the simplex into regions of feasible combinations of "true" beliefs given each report, for $k = 3$ and $n = 6$.

2.2 Applications and Extensions

2.2.1 Best responding to beliefs

One of the more common uses for elicited probabilities is testing whether subjects are behaving rationally, that is: are they best-responding to their beliefs? Having identified the bounds within which the true probability lies, the

Fig. 2 Reported and consistent true beliefs for $k = 3$ and $n = 6$



question remains as to which precise probabilities to use for further analysis. The midpoint (or centroid for $k \geq 3$) of the set of potentially underlying beliefs, or alternatively the stated frequencies, can be used as estimate of the underlying beliefs and as input to determine an estimate of expected utility. However, for statistical analysis of whether subjects are behaving rational it is advisable to work directly with the inferred set of possible beliefs.

When eliciting beliefs with the frequency guessing method, there is evidence that is consistent with the subject best-responding if the choice of action is payoff-maximising for *all* combinations of probabilities consistent with their reported beliefs. On the other hand,a subject can be identified as not being a best-responder if the action is not payoff-maximising for *any* of the probabilities that make their report optimal. Calculation of maximum and minimum expected values for a given strategy and set of stated beliefs is a straightforward constrained optimization problem, as presented in the following section.

As the number n of subjects belonging to the same role increases, the difference between estimated probabilities and probabilities that are consistent with the reported frequencies gets smaller. Thus, insights derived based on estimates for larger n more likely carry over to the rigorous treatment of reports.

Table 1 Maximum and minimum means and variances for $k = 3$ and $n = 10$

Stated belief	Max Mean	Min Mean	Δ Mean	Max Var	Min Var	Δ Var
(3,3,4)	0.727	0.667	0.061	0.083	0.070	0.013
(2,3,5)	0.788	0.722	0.066	0.077	0.062	0.015
(1,1,8)	0.909	0.833	0.076	0.068	0.042	0.026
(0,0,10)	1	0.917	0.083	0.039	0	0.039

2.2.2 Eliciting mean and variance

The bounds on probabilities derived in the previous subsection can be used to also place bounds on beliefs about the mean and variance of a distribution.

Assume that $s_i \in \mathbb{R}$ for $i = 1, .., k$. If one has elicited $\{b_i\}_{i=1}^k$ then one can derive 100% confidence intervals for functions of Y, such as EY and $VarY$.

Consider for instance EY. We obtain $EY \in [L(b), U(b)]$ where

$$L(b) = \min \left\{ \sum_{i=1}^{k} p_i s_i \text{ s.t. } p_i \geq 0, \sum p_i = 1, \text{ and } b_j p_i \leq (b_i + 1) p_j \forall i \neq j \right\}$$

and

$$U(b) = \max \left\{ \sum_{i=1}^{k} p_i s_i \text{ s.t. } p_i \geq 0, \sum p_i = 1, \text{ and } b_j p_i \leq (b_i + 1) p_j \forall i \neq j \right\}.$$

The width of this interval, $U(b) - L(b)$, depends on the precise values of the stated beliefs. Bounds are tight given that the underlying constraints as stated in Proposition 1 are necessary and sufficient. [4] Table 1 shows bounds on beliefs about means for $k = 3$, $n = 10$, and a selection of stated beliefs when $s_i = \frac{i}{3}$ for $i \in \{1, 2, 3\}$. Bounds on beliefs about variances can be calculated in a similar way and are shown in the same table.

Precision is decreasing in k and increasing in n. For example for $k = 5$ and $n = 10$ the maximal interval width when eliciting information about a mean is 0.138, whereas for $k = 3$ and $n = 20$ it is 0.045. For $k = 5$ and $n = 20$ the maximal interval width is 0.083.

3 Eliciting the Median and Quantiles

3.1 Median

In the following we present a method, the *median guessing method*, for eliciting the median of the belief distribution of a subject. The method determines when to award a prize based on the report of the subject and on n realizations of the underlying random variable Y.

[4] The same calculations also give the maximum and minimum possible expected utility of a subject playing a particular strategy when s_i is the payoff to the subject associated with that strategy when their partner plays strategy i.

We consider only the case where n is even. While all results extend in a straightforward way to odd n, in this case the instructions for subjects become less intuitive. For even n the instructions would be along the following lines: “Please tell us a number such that you think half the outcomes will be below or equal to that number. If your guess is correct you will receive R.” Odd n requires explaining that the prize will be paid if either $\frac{n-1}{2}$ or $\frac{n+1}{2}$ outcomes are weakly less than the guess. There is no advantage to this increased complexity[5] so when the maximal number of subjects in the role of the subject's partner is odd we suggest to drop one of the realizations at random.

Recall that the a real number $m_{1/2}$ is called a *median* of the random variable Y if $P\left(Y \leq m_{1/2}\right) = P\left(Y \geq m_{1/2}\right) = 1/2$. There are several different ways do adjust the definition when there is no real number $m_{1/2}$ with this property. In the following we define the median of Y equal to $m_{1/2}$ if $m_{1/2}$ is the smallest number m that satisfies $P\left(Y \leq m\right) \geq 1/2$. With this definition, the median both always exists and is unique due to the right continuity of any cdf.

According to the median guessing method, the subject is asked to report $m \in \mathbb{R}$ and is awarded a prize with value R if and only if half the realizations are smaller than or equal to m. Thus the prize is awarded with probability

$$f\left(m\right) = \binom{n}{n/2} P\left(Y \leq m\right)^{n/2} P\left(Y > m\right)^{n/2}.$$

Analogous to Section 2 without loss of generality we now investigate which reports maximize the probability of winning the prize. Again the only assumption on risk preferences is that $u\left(R\right) > u\left(0\right)$.

Consider first the case where the underlying distribution has no point masses.

Proposition 3 *Assume that n is even and $P\left(Y = y\right) = 0$ for all y. Then m maximizes f if and only if m is the median of Y.*

Proof Note that

$$f\left(m\right) = \binom{n}{n/2} \left(P\left(Y \leq m\right) \cdot \left(1 - P\left(Y \leq m\right)\right)\right)^{n/2} \tag{6}$$

Since $x\left(1-x\right)$ is maximized if and only if $x = 1/2$ we obtain that f is maximized if and only if $P\left(Y \leq m\right) = 1/2$ which completes the proof.

Consider now the case where point masses are possible. In particular, assume that Y only takes values on a finite grid $X = \{x_1, .., x_K\} \subset \mathbb{R}$ with $x_i < x_{i+1}$ for all $i = 1, .., K-1$. In this case it is enough to ask for reports that belong to X. Assume that each outcome occurs with strictly positive probability. We find that the report that maximizes the probability of receiving the prize is either equal to the median of their beliefs or equal to the next lower outcome.

[5] Unlike the methods in the previous section increasing n here does not increase precision.

Proposition 4 *Assume n even, $Y \in X$ with X finite and $P(Y = x) > 0$ for all $x \in X$. If $x_m \in X$ maximizes f then the median of Y is contained in $\{x_{m-1}, x_m\}$.*

Proof Following the proof of Proposition 4, f is maximized by x_m if and only if $\left|P(Y \leq x_m) - \frac{1}{2}\right| \leq \left|P(Y \leq x) - \frac{1}{2}\right|$ for all $x \in X$. Let x_i be the median of Y. If $P(Y \leq x_i) = 1/2$ then $x_m = x_i$. If $P(Y \leq x_i) > 1/2$ then $x_m \in \{x_{i-1}, x_i\}$ as $P(Y \leq x_{i-1}) < 1/2$.

It is easy to see that the proof also goes through if one allows for multiple medians by defining $m_{1/2}$ to be a median of Y if $P\left(Y \geq m_{1/2}\right) \geq 1/2$ and $P\left(Y \leq m_{1/2}\right) \geq 1/2$. One only has to adjust the statement of the proposition, replacing "the median" by "a median".

We suggest to use the report x_m as an estimate of the median, and to use the pair $\{x_{m-1}, x_m\}$ for any statistical analysis.

3.2 Quantiles

Our results for median elicitation are easily extended to general quantiles. Given $q \in (0,1)$ assume that we are interested in the quantile m_q of Y, where m_q is the smallest number that satisfies $P(Y \leq m_q) \geq q$. We show how to do this when nq is an integer. The idea is to award the prize if and only if the fraction q of the observations are smaller than or equal to m. The proofs follow analogously to those for the median, as $x^{q_1}(1-x)^{q_2}$ is maximized over all x for given $q_1, q_2 \in \mathbb{N}$ if and only if $x = \frac{q_1}{q_1+q_2}$.

4 Discussion

We have already discussed in some detail the primary advantages of these methods: validity for all non-trivial utility functions, and simplicity for subjects. The loss of precision in the elicitation of probabilities has also been mentioned.

One advantage of these methods is that the ease with which they can be communicated to subjects means that one can obtain measures of beliefs without adding a great deal of time or complexity to an experiment. Alternative methods typically require lengthy instructions, explanation of a formula or payoff tables, and in some cases additional tasks to approximate utility functions.

Which method to apply depends on the purpose of the elicitation, and the nature of the random variable one is considering. In any application involving expected utility theory the elicitation of probabilities and means will be necessary. However, if a measure of beliefs about a central tendency is desired, one must consider the tradeoff between the number of choices available and precision. For example, in a public goods game with an endowment of \$10 one could allow subjects to contribute only the whole endowment or nothing, or only multiples of \$5, thus reducing k to two or three.[6] For larger choice sets or continuous variables, the tests in Section 3 allow for precise estimation of medians and other quantiles, even with small numbers of subjects.

One theoretical issue that should be addressed is the assumption that the subject believes the realisations of Y to be independent, which is crucial for our results. However, statistical analysis of experimental data routinely rests upon this assumption, so we are confident that this will not be seen as a serious weakness. Furthermore, we do not require that the realisations are genuinely independent, as required for the validity of many statistical tests, but only that they are perceived to be independent by the subjects. While the assumption of independence should be borne in mind as a requirement for the validity of our results, we believe that in general it will be of little practical importance.

The assumption in the median elicitation described in Proposition 4 that every outcome is believed to have a non-zero probability of occurring could also be important. In most cases it is unlikely that subjects will see any outcome as *completely* impossible and would place some, perhaps infinitesimal, weight on it occurring. If there are outcomes that could credibly be seen as impossible however, this issue can be dealt with by simply removing these from the set of possible realizations prior to the elicitation.

One potential drawback is the possibility of weak incentives when eliciting probabilities with large n. For example, for $k = 3$, $n = 21$ and some sets of beliefs, the payment probability may be as low as 3.8%. This means P would have to be very large for the expected payoff to be of a significant size. It is an empirical question as to whether subjects will perceive the payment probability to be so low; any optimism about the accuracy of one's own predictions would lead to an increased perception of the payment probability. Also, the all or nothing nature of the payment should focus the attention of subjects, especially in comparison to the QSR where one can often guarantee half the maximum payoff by placing equal weight on all options. In the design of an experiment one may want to consider this trade-off between precision and strength of incentives when choosing n.

We add some more notes [as paragraph above already does some comparison] on the strength of incentives with the frequencing guessing method as compared to the QSR. From an ex-ante perspective QSR has a very flat payoff-function around the true belief, resulting in weak incentives to be precise. Similarly, the probability of winning the prize of the frequency guessing method may be of same magnitude as the optimal report for reports nearby. However, incentives are very different from an ex-post perspective, considering how other reports would have fared for the given realization. For a given realization, similar reports will generate similar payoffs under QSR. In contrast, under our method the prize is only awarded when reporting what had occurred. Thus, from an ex-post perspective the incentives of our method are dramatically stronger.

[6] Internet experiments would allow for very large n, and thus reasonably precise elicitation of beliefs about a larger choice set.

Weak incentives are not a problem in eliciting medians, however, as one can increase the probability of receiving the prize by reducing the number of realizations the payment is based on. In contrast to the case of probability elicitation, here reducing n does not result in loss of precision.

A related practical issue is that implementation of these methods could result in a large variance in payoffs between subjects, and also the total payment required for a session, which is often undesirable. With a small payment probability offset by a large prize the experimenter could be at risk of going over budget if too many subjects' predictions turn out to be accurate. This concern is easily circumvented by choosing a moderate prize, understanding that the utility of winning the prize incorporates non monetary pleasure of winning.

5 Conclusion

In this paper we have presented and characterized methods of belief elicitation which are extremely transparent to subjects and not dependent on restrictive assumptions about utility functions. We encourage experimentalists to use these methods in their own work, and especially to compare their empirical performance with other existing scoring rules.

References

Bhatt M, Camerer CF (2005) Self-referential thinking and equilibrium as states of mind in games: fmri evidence. Games and Economic Behavior 52:424–459

Blanco, Engelmann D, Koch AK, Norman HT (2010) Belief elicitation in experiments: is there a hedging problem? Experimental Economics 13(4):412–438

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

Cosmides L, Tooby J (1996) Are humans good intuitive statisticians after all? rethinking some conclusions from the literature on judgement under uncertainty. Cognition 58:1–73

Costa-Gomes MA, Weizsacker G (2008) Stated beliefs and play in normal-form games. The Review of Economic Studies 75:729–762

Gigerenzer G, Hoffrage U (1995) How to improve bayesian reasoning without instruction: Frequency formats. Psychological Review 102(4):684–704

Kahneman D, Tversky A (1983) Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. Psychological Review 90(4):293–315

Lipkus IM, Samsa G, Rimer BK (2001) General performance on a numeracy scale among highly educated samples. Medical Decision Making 21:37–44

Manski CF (2004) Measuring expectations. Econometrica 72(5):1329–1376

Offerman T, Sonnemans J, van de Kuilen G, Wakker PP (2009) A truth serum for non-bayesians: Correcting proper scoring rules for risk attitudes. Review of Economic Studies 76:1461–1489

Schapira MM, Nattinger AB, McHorney CA (2001) Frequency or probability? a qualitative study of risk communication formats used in health care. Medical Decision Making 21:459–467

Schlag KH, van der Weele J (2012) Eliciting probabilities, means, medians, variances and covariances without assuming risk neutrality?, mimeo, University of Frankfurt

Schwarz LM, Woloshin S, Black WC, Welch HG (1997) The role of numeracy in understanding the benefit of screening mammography. Annals of Internal Medicine 127:966–971

Wilcox NT, Feltovich N (2000) Thinking like a game theorist: Comment, mimeo, University of Houston