The Journal of Prediction Markets 2016 Vol 10 No 2 pp 14-21

A NOTE ON SANDRONI-SHMAYA BELIEF ELICITATION MECHANISM

Arthur Carvalho
Farmer School of Business,
Miami University
Oxford, OH,
USA, 45056
arthur.carvalho@miamioh.edu

ABSTRACT

Incentive-compatible methods for eliciting beliefs, such as proper scoring rules, often rely on strong assumptions about how humans behave when making decisions under risk and uncertainty. For example, standard proper scoring rules assume that humans are risk neutral, an assumption that is often violated in practice. Under such an assumption, proper scoring rules induce honest reporting of beliefs, in a sense that experts maximize their expected scores from a proper scoring rule by honestly reporting their beliefs.

Sandroni and Shmaya [*Economic Theory Bulletin*, volume 1, issue 1, 2013] suggested a remarkable mechanism based on proper scoring rules that induces honest reporting of beliefs without any assumptions on experts' risk attitudes. In particular, the authors claimed that the mechanism relies only on the natural assumptions of probabilistic sophistication and dominance. We suggest in this paper that the reduction of compound lotteries axiom is another assumption required for Sandroni and Shmaya's mechanism to induce honest reporting of beliefs. We further elaborate on the implications of such an extra assumption in light of recent findings regarding the reduction of compound lotteries axiom.

Keywords: Proper Scoring Rules; Belief Elicitation; Reduction of Compound Lotteries Axiom

1 INTRODUCTION

Consider the scenario where a decision maker is interested in a forecast (*belief*), which is represented by a discrete probability distribution over a set of exhaustive and mutually exclusive outcomes $\theta_1, \theta_2, \ldots, \theta_n$. The decision maker elicits beliefs from human experts, who have no influence on or stakes in the outcomes of interest. We denote an expert's belief by the probability vector $\boldsymbol{p} = (p_1, p_2, \ldots, p_n)$, where p_k is his subjective probability regarding the occurrence of outcome θ_k. Experts are potentially strategic, meaning that they are not necessarily honest when reporting their beliefs. Therefore, we distinguish between an expert's *true belief* $\boldsymbol{p}$ and his *reported belief* $\boldsymbol{q} = (q_1, q_2, \dots, q_n)$. Clearly, from a decision making perspective, it is desirable to obtain $\boldsymbol{q} = \boldsymbol{p}$. When this happens, we say that the expert is honestly reporting his belief.

Proper scoring rules (Winkler & Murphy 1968) are traditional techniques to induce risk-neutral agents to honestly report their beliefs. A *scoring rule* $R(\boldsymbol{q}, \theta_x)$ evaluates the accuracy of a reported belief $\boldsymbol{q}$ by providing a real-valued score upon observing an outcome θ_x, for $x \in \{1, \dots, n\}$. In order to become relevant, scores are often coupled with financial and/or social-psychological rewards, which naturally implies that experts seek to maximize the obtained scores. A scoring rule is called (*strictly*) *proper* when an expert receives his maximum expected score if (and only if) his reported belief $\boldsymbol{q}$ matches his true belief $\boldsymbol{p}$ (Winkler & Murphy 1968). The *expected score* of $\boldsymbol{q}$ at $\boldsymbol{p}$ for a real-valued scoring rule $R(\boldsymbol{q}, \theta_x)$ is: $\boldsymbol{E_p}[R(\boldsymbol{q}, \cdot)] = \sum_{x=1}^{n} p_x R(\boldsymbol{q}, \theta_x)$. Arguably, the logarithmic proper scoring rule, $R(\boldsymbol{q}, \theta_x) = \log q_x$, and the quadratic rule, $R(\boldsymbol{q}, \theta_x) = 2q_x - \sum_{k=1}^{n} q_k^2$, are the most popular proper scoring rules.

Proper scoring rules rely on the assumption that experts are risk neutral, which is a strong and often unrealistic assumption when experts are humans (Weber & Chapman 2005; Armantier & Treich 2013). An expert's risk attitude might influence the way the expert reports his belief under standard proper scoring rules. For example, risk-seeking experts tend to report sharp beliefs, whereas risk-averse experts tend to report beliefs close to the uniform distribution (Winkler & Murphy 1970; Armantier & Treich 2013; Holt & Laury 2002; Offerman et al. 2009).

An alternative when risk neutrality does not hold true is to reward the expert using the scoring rule $U^{-1}(R(\boldsymbol{q}, \theta_x))$, which results into a proper scoring rule for the utility function $U(\cdot)$ (Winkler 1969). Clearly, this approach relies on two conditions: 1) the decision maker knows that the expert behaves according to expected utility theory; and 2) the decision maker knows the shape of the expert's utility function $U(\cdot)$. When utility functions are unknown, standard proper scoring rules for eliciting the probability of an event using deterministic payments no longer exist (Schlag & van der Weele 2013).

One approach to circumvent the above impossibility result is to elicit the components that drive an expert's risk attitude towards uncertainty before eliciting the expert's belief using a proper scoring rule. Thereafter, the decision maker is able to calibrate the expert's reported belief *a posteriori* by removing the influence of those components in order to obtain the expert's true belief (Carvalho 2015; Offerman et al. 2009; Kothiyal et al. 2011). Clearly, the decision maker must assume that experts behave according to a certain decision model in order for this approach to work. When the assumed model is wrong, the final calibrated belief can be very different from the expert's true belief (Carvalho 2015).

Another alternative to circumvent the above impossibility result is to make payments in lotteries, instead of using deterministic payments. For example, Allen (1987) presented a randomized payment method that relies on the linearization of utility through conditional lottery tickets to induce honest reporting when an expert's utility function is unknown. More recently, Karni (2009) proposed a method with two fixed prizes where the payment function is determined by comparing the expert's reported probability value to a random number drawn uniformly from [0,1]. Under Karni's method, if an expert exhibits probabilistic sophistication and dominance, then it is in the best interest of the expert to report honestly regardless of his risk attitudes.

In spirit, the methods by Allen (1987) and Karni (2009) are analogous to the classic Becker–DeGroot–Marschak mechanism (Becker et al. 1964). Some subjects have been found to have a hard time dealing with Becker-DeGroot-Marschak mechanisms in experimental settings (Cason & Plott 2014; Plott & Zeiler 2005; Rutström 1998). Sandroni & Shmaya (2013) proposed a simpler, yet elegant stochastic payment scheme based on proper scoring rules, which we discuss next. We argue, however, that Sandroni and Shmaya's mechanism only induces honest reporting of beliefs when the reduction of compound lotteries axiom holds true, an assumption not mentioned by the authors. We discuss in Section 3 the practical implications of such an assumption.

2 SANDRONI-SHMAYA MECHANISM

From an expert's perspective, reporting a belief under a proper scoring rule is equivalent to choosing a lottery over a potentially infinite number of lotteries (Carvalho 2015). Consider a scenario involving only two outcomes ($n = 2$). Formally, the implicitly chosen lottery associated with a reported belief $\boldsymbol{q} = (q_1, q_2)$ is:

$$[R(\boldsymbol{q}, \theta_1): p_1;\ R(\boldsymbol{q}, \theta_2): p_2]$$

which means that the expert believes he will receive the score $R(\boldsymbol{q}, \theta_1)$ with probability p_1, and $R(\boldsymbol{q}, \theta_2)$ otherwise. Now, consider the following lotteries involving the fixed values $x_{max} > x_{min}$ and probabilities $0 \leq \rho, \rho' \leq 1$:

$$A = [x_{min}: \rho;\ x_{max}: 1-\rho] \text{ and } B = [x_{min}: \rho';\ x_{max}: 1-\rho']$$

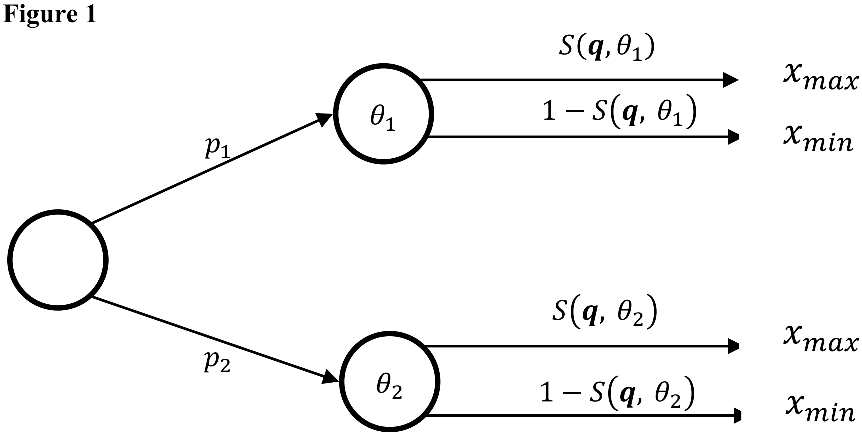
We say that *probabilistic dominance* holds true when an expert strictly prefers A over B if and only if $\rho < \rho'$. Sandroni & Shmaya (2013) argued that probabilistic dominance is the only assumption required in our setting (where probabilistic sophistication already holds true) when inducing honest reporting of beliefs. Specifically, the authors proposed the following payment scheme:

If outcome θ_1 happens: the expert receives the payoff x_{max} with probability $S(\boldsymbol{q}, \theta_1)$ and x_{min} with probability $1 - S(\boldsymbol{q}, \theta_1)$;

If outcome θ_2 happens: the expert receives the payoff x_{max} with probability $S(\boldsymbol{q}, \theta_2)$ and x_{min} with probability $1 - S(\boldsymbol{q}, \theta_2)$;

where $S(\boldsymbol{q}, \theta_x) \in [0,1]$ is a normalized proper scoring rule. We note that any bounded proper scoring rule can be used to create a normalized scoring rule, which in turn is still proper since a positive affine transformation of a proper scoring rule is still proper (Gneiting & Raftery 2007). We argue that, from an expert's point of view, the above payment scheme is equivalent to the following compound lottery:

Figure 1



Sandroni & Shmaya (2013) suggested that the probability of receiving the highest payment x_{max} is:

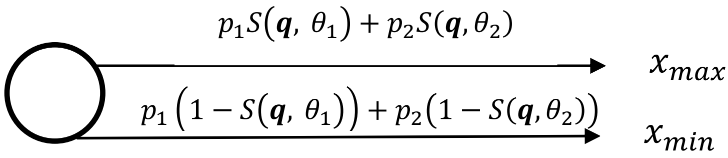
$$p_1 * S(\boldsymbol{q}, \theta_1) + p_2 * S(\boldsymbol{q}, \theta_2)$$

Furthermore, under the assumption of probabilistic dominance, the expert will behave so as to maximize the above probability. Note that the above value is the expected score of the proper scoring rule $S(\cdot)$, which in turn is maximized when $\boldsymbol{q} = \boldsymbol{p}$, *i.e.*, when the expert reports honestly.

A remarkable characteristic of Sandroni and Shmaya's mechanism is that it does not rely on any assumptions about experts' risk attitudes. By fixing a couple of prizes, all bounded proper scoring rules can be transformed into randomized rules that induce honest reporting for all risk preferences. This mechanism differs from Becker–DeGroot–Marschak based mechanisms in that no external randomization device other than nature is required to determine an expert's payment. One implicit assumption not mentioned by

Sandroni & Shmaya (2013) is, however, that the expert must be indifferent between the lottery in Figure (1) and the following lottery:

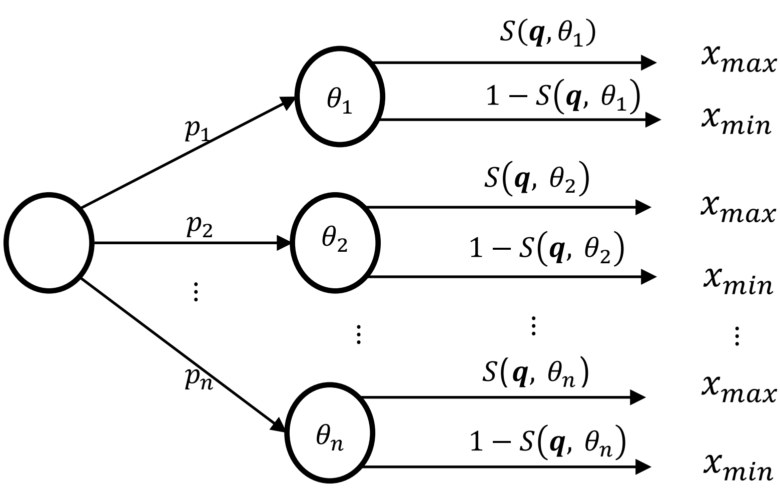
Figure 2



In other words, there is an implicit assumption that the *reduction of compound lotteries* (ROCL) axiom holds true (Harrison et al. 2015).

The original mechanism by Sandroni & Shmaya (2013) focused only on binary outcomes. The authors mentioned that the mechanism "*can be extended to many events and many experts, by running separate mechanisms to different experts and events*". Clearly, this approach is problematic when the number of outcomes *n* is large. This happens because the decision maker has to run the mechanism once for eliciting the probability associated with each individual outcome. We note next that the same mechanism is easily extendable to multiple outcomes, where the payment scheme becomes:

Figure 3



- *If outcome θ_1 happens*: the expert receives the payoff x_{max} with probability $S(\boldsymbol{q}, \theta_1)$ and x_{min} with probability $1 - S(\boldsymbol{q}, \theta_1)$;
- *If outcome θ_2 happens*: the expert receives the payoff x_{max} with probability $S(\boldsymbol{q}, \theta_2)$ and x_{min} with probability $1 - S(\boldsymbol{q}, \theta_2)$;
- …

- *If outcome* θ_n *happens*: the expert receives the payoff x_{max} with probability $S(\boldsymbol{q}, \theta_n)$ and x_{min} with probability $1 - S(\boldsymbol{q}, \theta_n)$;

The above payment scheme is equivalent to the compound lottery in Figure 3.

Under the ROCL assumption, the above lottery reduces to:

$$\left[x_{min}: \sum_{x=1}^{n} p_x\big(1 - S(\boldsymbol{q}, \theta_x)\big);\ x_{max}: \sum_{x=1}^{n} p_x S(\boldsymbol{q}, \theta_x)\right],$$

in which case the probability of receiving the highest payoff is maximized when $\boldsymbol{q} = \boldsymbol{p}$. Once again the crucial assumption missing in the analysis by Sandroni & Shmaya (2013) is that the ROCL axiom must hold true. We discuss in the following section the validity of such an assumption.

3 DISCUSSION

At this point, a question that arises is: when does the ROCL axiom hold true in practice? Results from prior empirical work are mostly negative in terms of the validity of the ROCL assumption, *e.g.*, see Appendix B in the paper by Harrison et al. (2015). However, Harrison et al. (2015) suggested that these negative results could had been driven by the underlying payment scheme. In particular, the use of the well-known random lottery incentive mechanism might induce violations of ROCL. The random lottery incentive mechanism repeatedly (for a total of k times) presents two lotteries to subjects and asks them to select the most desirable one. Eventually, the mechanism randomly selects and plays out one of the chosen lotteries. Interesting, Harrison et al. (2015) showed that there is no longer evidence of violations of ROCL when subjects face only one binary choice ($k = 1$).

Experts also face a single-decision situation when they report their beliefs under Sandroni and Shmaya's mechanism as defined is this paper. Hence, ROCL might be satisfied according to the results by Harrison et al. (2015) and, consequently, Sandroni and Shmaya's mechanism might still induce honest reporting of beliefs. We note, however, that the setting investigated by Harrison et al. (2015) involved only binary choices, *i.e.*, choices over two lotteries. When an expert is asked to report a belief under a proper scoring rule, each potential reported belief generates a lottery. Hence, the expert's choice is over a potentially infinite number of lotteries (reported beliefs). Furthermore, Harrison et al. (2015) derived their results in the domain of risk, which is a subcase of the uncertainty domain when all the experts' beliefs are equal to each other (Wakker 2010). The extent to which the results by Harrison et al. (2015) hold true under uncertainty and for an infinite number of choices is still an open and exciting research question. An answer to such a question will determine the empirical validity of Sandroni and Shmaya's mechanism for inducing honest reporting of beliefs.

4 REFERENCES

Allen, F., 1987. Discovering Personal Probabilities when Utility Functions are Unknown. *Management Science*, 33(4), pp.542–544.

Armantier, O. & Treich, N., 2013. Eliciting Beliefs: Proper Scoring Rules, Incentives, Stakes and Hedging. *European Economic Review*, 62, pp.17–40.

Becker, G.M., Degroot, M.H. & Marschak, J., 1964. Measuring Utility by a Single-Response Sequential Method. *Behavioral Science*, 9(3), pp.226–232.

Carvalho, A., 2015. Tailored Proper Scoring Rules Elicit Decision Weights. *Judgment and Decision Making*, 10(1), pp.86–96.

Cason, T. & Plott, C., 2014. Misconceptions and Game Form Recognition: Challenges to Theories of Revealed Preference and Framing. *Journal of Political Economy*, 122(6), pp.1235–1270.

Gneiting, T. & Raftery, A.E., 2007. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477), pp.359–378.

Harrison, G.W., Martínez-Correa, J. & Swarthout, J.T., 2015. Reduction of Compound Lotteries with Objective Probabilities: Theory and Evidence. *Journal of Economic Behavior and Organization*, 119, pp.32–55.

Holt, C.A. & Laury, S.K., 2002. Risk Aversion and Incentive Effects. *American Economic Review*, 92(5), pp.1644–1655.

Karni, E., 2009. A Mechanism for Eliciting Probabilities. *Econometrica*, 77(2), pp.603–606.

Kothiyal, A., Spinu, V. & Wakker, P., 2011. Comonotonic Proper Scoring Rules to Measure Ambiguity and Subjective Beliefs. *Journal of Multi-Criteria Decision Analysis*, 17(3-4), pp.101–113.

Offerman, T., Sonnemans, J., Van De Kuilen, G., Wakker, P.P., 2009. A Truth Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes. *Review of Economic Studies*, 76(4), pp.1461–1489.

Plott, C.R. & Zeiler, K., 2005. The Willingness to Pay-Willingness to Accept Gap, the "Endowment Effect," Subject Misconceptions, and Experimental Procedures for Eliciting Valuations. *American Economic Review*, 95(3), pp.530–545.

Rutström, E.E., 1998. Home-Grown Values and Incentive Compatible Auction Design. *International Journal of Game Theory*, 27(3), pp.427–441.

Sandroni, A. & Shmaya, E., 2013. Eliciting Beliefs by Paying in Chance. *Economic Theory Bulletin*, 1(1), pp.33–37.

Schlag, K.H. & van der Weele, J.J., 2013. Eliciting Probabilities, Means, Medians, Variances and Covariances without Assuming Risk Neutrality.

Theoretical Economics Letters, 3(1), pp.38–42.

Wakker, P.P., 2010. *Prospect theory: For Risk and Ambiguity*, Cambridge University Press.

Weber, B.J. & Chapman, G.B., 2005. Playing for Peanuts: Why is Risk Seeking More Common for Low-Stakes Gambles? *Organizational Behavior and Human Decision Processes*, 97(1), pp.31–46.

Winkler, R.L., 1969. Scoring Rules and the Evaluation of Probability Assessors. *Journal of the American Statistical Association*, 64(327), pp.1073–1078.

Winkler, R.L. & Murphy, A.H., 1968. “Good” Probability Assessors. *Journal of Applied Meteorology*, 7(5), pp.751–758.

Winkler, R.L. & Murphy, A.H., 1970. Nonlinear Utility and the Probability Score. *Journal of Applied Meteorology*, 9, pp.143–148.