Decision Scoring Rules

A precursor of this paper was published in WINE'20.

Caspar Oesterheld (cpo11@duke.edu) and Vincent Conitzer
Duke University, Department of Computer Science

Abstract

We consider a setting in which a principal faces a decision and asks an external expert for a recommendation as well as a probabilistic prediction about what outcomes might occur if the recommendation were implemented. The principal then follows the recommendation and observes an outcome. Finally, the principal pays the expert based on the prediction and the outcome, according to some *decision scoring rule*. In this paper, we ask the question: What does the class of *proper* decision scoring rules look like, i.e., what scoring rules incentivize the expert to honestly reveal both the action he believes to be best for the principal and the prediction for that action? We first show that in addition to an honest recommendation, proper decision scoring rules can only incentivize the expert to reveal the expected utility of taking the recommended action. The principal cannot strictly incentivize honest reports on other aspects of the conditional distribution over outcomes without setting poor incentives on the recommendation itself. We then characterize proper decision scoring rules as ones which give or sell the expert shares in the principal's project. Each share pays, e.g., \$1 per unit of utility obtained by the principal. Owning these shares makes the expert want to maximize the principal's utility by giving the best-possible recommendation. Furthermore, if shares are offered at a continuum of prices, this makes the expert reveal the value of a share and therefore the expected utility of the principal conditional on following the recommendation. We extend our analysis to eliciting recommendations and predictions from multiple experts. With a few modifications, the above characterization for the single-expert case carries over. Among other implications, this characterization implies that in generic mechanisms no expert should be able to "short-sell" shares in the principal's project and thereby profit if the project goes poorly.

1 Introduction

Consider a firm that is about to make a major strategic decision. It wishes to maximize the expected value of the firm. It hires an expert to consult on the decision. The expert is better informed than the firm, but it is commonly understood that the outcome conditional on the chosen course of action is uncertain even for the expert. The firm can commit to a compensation package for the expert; compensation can be conditional both on the expert's predictions and on what happens (e.g., in terms of the value of the firm) after a decision is made. (The compensation cannot depend on what would have happened if another action had been chosen.) The firm cannot or does not want to commit to an arbitrary mapping from expert reports to actions: once the report is made, the firm will always choose the action that maximizes expected value, according to that report. What compensation schemes will incentivize the expert to report truthfully? One straightforward solution is to give the expert a fixed share of the firm at the outset. Are there other schemes that also reward accurate predictions? What compensation schemes are effective if the firm can consult multiple experts?

Our approach to formalizing and answering these questions is inspired by existing work on eliciting honest predictions about an event that the firm or *principal* cannot influence. In the single-expert case, such elicitation mechanisms are known as *proper scoring rules* (Brier, 1950; Good, 1952, Section 8; McCarthy, 1956; Savage, 1971; Gneiting and Raftery, 2007). Formally, a scoring rule for prediction s takes as input a probability distribution $\hat{P}$ reported by the expert, as well as the actual outcome ω, and assigns a *score* or *reward* $s(\hat{P}, \omega)$. A scoring rule s is proper if the expert maximizes his[1] expected score by reporting as $\hat{P}$ his true beliefs about how likely different outcomes are. The class of proper scoring rules has been characterized in prior work [e.g., Gneiting and Raftery, 2007, Section 2]. This characterization also provides a foundation for the design of proper scoring rules that are *optimal* with respect to a specific objective and potentially under additional constraints (Osband, 1989; Neyman, Noarov, and Weinberg, 2020; Hartline et al., 2020). Work on proper scoring rules has also contributed to work on eliciting information from multiple experts via so-called *prediction markets* [e.g., Hanson, 2003; Pennock and Sami, 2007]. For example, in a *market scoring rule*, agents successively update the probability estimate, and an agent that updated the estimate from $\hat{P}_t$ to $\hat{P}_{t+1}$ is eventually rewarded $s(\hat{P}_{t+1}, \omega) - s(\hat{P}_t, \omega)$. Alternative designs, which resemble real-world securities markets, let experts trade *Arrow-Debreu securities* that each pay out a fixed amount – say, \$1 – if a given event happens, and \$0 otherwise. Then, at any point, the price at which this security trades can be seen as the current market consensus of the probability that the event takes place. There is a close correspondence between Arrow-Debreu securities markets and market scoring rules (Hanson, 2003; Hanson, 2007; Pennock and Sami, 2007, Section 4; Chen and Pennock, 2007; Agrawal et al., 2009; Chen and Vaughan, 2010).

Contributions In this paper, we derive a similar characterization of what we call *proper decision scoring rules* – scoring rules that incentivize the expert to honestly report the best available action in addition to making an honest prediction. We introduce our setup and the concept of propriety in detail in Section 2. We show that proper decision scoring rules cannot give the expert *strict* incentives to report any properties of the outcome distribution under the recommended action other than its expected utility (Section 3). Intuitively, rewarding the expert for getting anything else about the distribution right will make him recommend actions whose outcome is easy to predict as opposed to actions with high expected utility. Hence, the expert's reward can depend only on the reported expected utility for the recommended action. Next we show that the scoring rule must be affine. Using these results, we then obtain four characterizations of proper decision scoring rules (Section 5), two of which are analogous to existing results on proper affine scoring, including the characterizations of proper scoring rules by Gneiting and Raftery (2007) and the generic characterization of proper affine scoring rules by Frongillo and Kash (2014). One of the other two characterizations (Theorem 5.6, which we have not seen anywhere in this form for affine scoring), has an especially intuitive interpretation: the principal offers shares in her project to the expert at some pricing schedule. The price schedule does not depend on the action chosen. Thus, given the chosen action, the expert is incentivized to buy shares up to the point where the price of a share exceeds the expected value of the share, thereby revealing the principal's expected utility. Moreover, once the expert has some positive share in the principal's utility, he will be (strictly) incentivized to recommend an optimal action. In Section 7, we discuss the implications of our characterization for mechanism for eliciting decision-relevant mechanisms from multiple experts. Finally, we discuss related work in Section 8.

2 Setup

Throughout most of this paper, we consider the following setup.

A *principal* faces a choice from a finite set A of at least two actions. Decisions stochastically give rise to outcomes from a finite set Ω. The principal would like to choose an action that maximizes the expectation of a utility function $u \colon \Omega \to \mathbb{R}$. Before making a choice, she privately observes the value of an evidence variable E, which has values in some finite, non-empty set H. The principal

[1] Following convention, the principal is grammatically female (pronouns "she/her/hers") and the expert is grammatically male (pronouns "he/him/his") throughout this paper.

knows neither the distributions over Ω arising from any $a \in A$, conditional on any $e \in H$, nor the distribution of E.

To make a choice, the principal consults an *expert*. The expert does have probabilistic beliefs. Specifically, he believes E to be distributed according to some $Q \in \Delta(H)$. Furthermore, for each $e \in H$ and $a \in A$, he believes that the outcome will be distributed according to $P(\cdot \mid e, a) \in \Delta(\Omega)$ if e is observed and action a is taken.

The principal asks the expert for a report $\alpha \in A^H$. We will call this the expert's *recommendation*. The principal's intention is for the expert to report α such that for all e, $\alpha(e)$ maximizes the principal's utility given the expert's belief P, i.e.,

$$\alpha(e) \in \arg\max_{a \in A} \mathbb{E}\left[u(O)\right] \coloneqq \sum_{\omega \in \Omega} P(\omega \mid a, e) u(\omega). \tag{1}$$

for all $e \in H$. We call such recommendations α *honest*. However, the expert can report any $\alpha \in A^H$; the principal is not able to directly determine whether a submitted recommendation is honest.

The principal further asks the experts to report a *conditional outcome prediction* $\hat{P}_\alpha \in \Delta(\Omega)^H$. The principal's intention is for the expert to report $\hat{P}_\alpha$ s.t. for all $e \in H, \omega \in \Omega$,

$$P_\alpha(\omega \mid e) = P(\omega \mid e, \alpha(e)).$$

We analogously call this the honest prediction for the recommendation α. Note that a prediction can only be honest relative to a recommendation and that honesty is defined even for dishonest recommendations. Again, the expert can submit any $\hat{P}_\alpha \in \Delta(\Omega)^H$ and the expert cannot immediately determine whether the prediction is honest.

Further, the principal asks the expert for an *evidence prediction* $\hat{Q} \in \Delta(H)$. The principal's intention is for the expert to honestly report $\hat{Q} = Q$, though again the principal cannot directly verify honesty.

Once the expert has submitted his *report* $(\alpha, \hat{P}_\alpha, Q)$ and the principal observes $E = e \in H$, the principal chooses the recommended action $\alpha(e)$. He then observes an outcome ω.

To incentivize the expert to report honestly, the principal rewards the expert using a *decision scoring rule* (DSR) $s \colon \Delta(H) \times \Delta(\Omega)^H \times H \times \Omega \to \mathbb{R}$, which maps the expert's evidence prediction $\hat{Q}$, conditional outcome prediction $\hat{P}_\alpha$, the true evidence e, and the true outcome ω onto a score $s(\hat{Q}, \hat{P}_\alpha, e, \omega)$.

The question we ask in this paper is what DSRs incentivize the expert to report honestly. We define this formally as follows.

Definition 1. We say that a DSR s is *proper* if for all beliefs $P(\cdot \mid \cdot, \cdot) \in \Delta(\Omega)^{H \times A}$ and all possible recommendations $\hat{\alpha} \in A^H$ and predictions $\hat{P}_\alpha \in \Delta(\Omega)^H$ we have

$$\mathbb{E}_{E \sim Q, O \sim P}\left[s(\hat{P}_\alpha, E, O) \mid E, \hat{\alpha}(E)\right] \leq \mathbb{E}_{E \sim Q, O \sim P}\left[s(Q, P_{\alpha^*}, E, O) \mid E, \alpha^*(E)\right]$$

for some honest recommendation α^*.

Our goal is to to characterize proper DSRs. However, while this propriety implies that the expert has no bad incentives, it does not require that the expert has any good incentives. For example, any constant s is proper. We might therefore be interested in the structure of *strictly* proper DSRs, i.e., ones where inequality 1 is strict unless $(\hat{\alpha}, Q, P_\alpha)$ is an honest report. As we will see (Lemma 3.3), no DSR is strictly proper in this sense. In the following we therefore define partially strict versions of propriety.

Definition 2. We say that a proper s is *strictly proper w.r.t. the recommendation* if for all beliefs $P(\cdot \mid \cdot, \cdot) \in \Delta(\Omega)^{H \times A}, Q \in \Delta(H)$ and all possible reports $(\alpha, \hat{Q}, \hat{P}_\alpha)$ with dishonest recommendation α, there exists an honest report $(\alpha^*, Q, P_{\alpha^*})$ s.t.

$$\mathbb{E}_{E \sim Q, O \sim P}\left[s(\hat{P}_\alpha, E, O) \mid E, \hat{\alpha}(E)\right] < \mathbb{E}_{E \sim Q, O \sim P}\left[s(P_{\alpha^*}, E, O) \mid E, \alpha^*(E)\right].$$

Example 1 (Linear scoring rules)**.** Let $\mathbf{c} \in \mathbb{R}^H_{\geq 0}$. Then $s\colon (Q, P_\alpha, e, \omega) \mapsto c_e u(\omega)$ is proper. If furthermore, $c_e > 0$ then s is strictly proper w.r.t. the recommendation for e.

A natural interpretation of this DSR is that the principal gives the experts fixed numbers of shares in her project that pay, say, \$1 per unit of utility obtained by the principal if a particular evidence e is observed.

Definition 3. We say that a proper s is *strictly proper w.r.t. the evidence distribution* if for all beliefs $P(\cdot \mid \cdot, \cdot) \in \Delta(\Omega)^{H \times A}$ and all possible reports $(\alpha, \hat{Q}, \hat{P}_\alpha)$ with $\hat{Q} \neq Q$, there exists an honest report $(\alpha^*, Q, P_{\alpha^*})$ s.t.

$$\mathbb{E}_{E \sim Q, O \sim P}\left[s(\hat{P}_\alpha, E, O) \mid E, \hat{\alpha}(E)\right] < \mathbb{E}_{E \sim Q, O \sim P}\left[s(P_{\alpha^*}, E, O) \mid E, \alpha^*(E)\right].$$

Example 2 (Brier's (1950) scoring rule for evidence prediction)**.** Consider the scoring rule

$$s(\hat{Q}, \hat{P}_\alpha, e, \omega) = 2\hat{Q}(e) - \sum_{e' \in H} \hat{Q}(e')^2.$$

This is Brier's quadratic scoring applied for scoring $\hat{Q}$ as a prediction of e. It can be shown that s defined in this way is proper and strictly proper w.r.t. the evidence prediction. (In fact any strictly proper scoring rule for prediction (as defined and characterized by, e.g., Gneiting and Raftery, 2007, Section 2) is strictly proper w.r.t. the evidence prediction when transferred in this way.) Of course, it is not strictly proper w.r.t. anything else, because the score does not depend on P_α, ω at all.

We could combine Examples 1 and 2 to obtain a DSR that is proper and strictly proper w.r.t. recommendations and evidence predictions. The next example illustrates the troubles that arise when we aim for an analogous strict propriety w.r.t. the outcome prediction. (As noted above, we will show in Section 3 that these troubles cannot be overcome.)

Example 3 (Misapplying Brier's (1950) scoring rule for outcome prediction)**.** For simplicity, imagine that $H = \{e\}$ and we use the following scoring rule

$$s(\hat{Q}, \hat{P}_\alpha, e, \omega) = 2\hat{P}_\alpha(\omega) - \sum_{\omega' \in \Omega} \hat{P}_\alpha(\omega')^2.$$

Once again, this is Brier's scoring rule. However, this time it is applied to the submitted outcome prediction. Again it can be shown that to maximize his expert score, the expert has to submit $\hat{P}_\alpha = P_\alpha$ honestly. Nevertheless, a DSR s defined in this way is *not* proper, because to maximize his expected score, the expert has to recommend an action $\alpha(e)$ s.t. the true distribution $P_{\alpha(e)}$ is easy to predict (in the sense of having a high Brier score under honest prediction). For instance, suppose that $\Omega = \{\omega_1, ..., \omega_m\}$, that the optimal action a^* leads to the uniform distribution $O_{a^*} = \frac{1}{m} * \omega_1 + ... + \frac{1}{m} * \omega_m$, while a' leads to a bad outcome deterministically, $O_{a'} = 1 * \omega_1$. Then the expert will (assuming $m > 1$) always prefer recommending the suboptimal a', since

$$\mathbb{E}\left[s(\hat{Q}, P_{a'}, E, O_{a'})\right] = 1 > \frac{1}{m} = \frac{2}{m} - \sum_{w' \in \Omega} \frac{1}{m^2} = \mathbb{E}\left[s(\hat{Q}, P_{a^*}, E, O_{a^*})\right].$$

Definition 4. We say that s is strictly proper w.r.t. the means if it is proper and for all beliefs P, Q and all reports $(\alpha, Q, \hat{P}_\alpha)$ with honest evidence prediction for which there is e with

$$\mathbb{E}_{O \sim \hat{P}_\alpha}\left[u(O) \mid e\right] \neq \mathbb{E}_{O \sim P}\left[u(O) \mid e, \alpha(e)\right],$$

the report (α, Q, P_α) with honest conditional outcome prediction P_α,

$$\mathbb{E}_{E \sim Q, O \sim P}\left[s(Q, \hat{P}_\alpha, E, O) \mid E, \hat{\alpha}(E)\right] < \mathbb{E}_{E \sim Q, O \sim P}\left[s(P_{\alpha^*}, E, O) \mid E, \alpha^*(E)\right].$$

3 Only means matter

Next, we prove that if a DSR is to be proper, it can only strictly incentivize the expert to be honest about the optimal (recommended) actions and the expected utility of those recommended actions. That is, proper scoring rules can be strictly proper w.r.t. the recommendation, the means and the evidence prediction, but nothing else. For example, proper scoring rules cannot strictly incentivize the expert to honestly reveal the variance of the utility given that the recommended action is taken. First, we need two simple lemmas.

Lemma 3.1. *If s is proper, then $\mathbb{E}_{E\sim Q, O\sim P_\alpha}[s(Q,P,E,O)]$ is continuous in P,Q in the set of P,Q with full support.*

Lemma 3.2. *Let s be a proper DSR and $P_\alpha, P'_{\alpha'} \in \Delta(\Omega)^H$ be s.t. for all $e \in H$*

$$\min_{\omega\in\Omega} u(\omega) < \mathbb{E}_{O\sim P_\alpha}[u(O) \mid e] = \mathbb{E}_{O\sim P'_{\alpha'}}[u(O) \mid e] < \max_{\omega\in\Omega} u(\omega). \tag{2}$$

Then

$$\mathbb{E}_{E\sim Q, O\sim P_\alpha}[s(Q,P_\alpha,E,O)] = \mathbb{E}_{E\sim Q, O\sim P_{\alpha'}}[s(Q,P_{\alpha'},E,O)]. \tag{3}$$

It is worth noting that the proof is based on the lack of "space" in the set $\mathbb{R}$ of possible scores. We could imagine experts who maximize a lexicographic score. Then our result only shows that the lexically highest value of the scores – under honest reporting – of two equally good recommendations must be the same. But the lexically lower values could be given according to some scoring rule for prediction (such as the quadratic scoring rule) and thus make the expert prefer one of two recommendations with equal expected utility for the expert.

Note also that the lemma only shows that the expected scores *across realizations of E* under honest report of $P_\alpha, P_{\alpha'}$ are the same. For individual realizations e, the expected scores can be different, as the following example shows.

Example 4. Define s as follows. Based on the reported $\hat{Q}, \hat{P}_\alpha$ a special e^* will be determined in a way described below. We then let $s(\hat{Q},\hat{P}_\alpha,e,\omega) = u(\omega)$ for $e \neq e^*$ and $s(\hat{Q},\hat{P}_\alpha,e^*,\omega) = 2u(\omega)$. In the "giving shares" interpretation, explained after Example 1, the expert receives a single share and a share that only pays for e^*. Let e^* be selected from $\arg\max_e \mathbb{E}_{O\sim\hat{P}_\alpha}[u(O) \mid e]\,\hat{Q}(e)$. Importantly, ties are broken based on $\hat{P}_\alpha$, for example, by the entropy of $\hat{P}_\alpha(\cdot \mid e)$. Note that such s is proper, because $\mathbb{E}_{O\sim P_\alpha}[u(O) \mid e]\,Q(e)$ is the value of an extra share for e to the expert. Under this proper DSR, two different, honestly reported α, P_α and $\alpha', P_{\alpha'}$ may then differ in their expected scores for e, e', if they both claim e, e' to both be in the above argmax but the tie is broken differently for P_α versus $P_{\alpha'}$.

We now get to the main result of the present section: aside from degenerate cases, the expert does not change his expected score by misreporting $\hat{P}_\alpha$ (relative to truthfully reporting P_α), as long as $\hat{P}_\alpha$ gives the accurate means and does not assign zero probability to any outcome that is in fact possible.

Lemma 3.3. *Let s be a proper DSR, $Q \in \Delta(H)$ and $P_\alpha, \hat{P}_\alpha \in \Delta(\Omega)^H$ be s.t. for all $e \in H$*

$$\min_{\omega\in\Omega} u(\omega) < \mathbb{E}_{O\sim P_\alpha}[u(O) \mid e] = \mathbb{E}_{O\sim \hat{P}_\alpha}[u(O) \mid e] < \max_{\omega\in\Omega} u(\omega) \tag{4}$$

and $\mathrm{supp}(P_\alpha(\cdot \mid e)) \subseteq \mathrm{supp}(\hat{P}_\alpha(\cdot \mid e))$. *Then*

$$\mathbb{E}_{E\sim Q, O\sim P_\alpha}[s(Q,P_\alpha,E,O)] = \mathbb{E}_{E\sim Q, O\sim P_\alpha}\left[s(Q,\hat{P}_\alpha,E,O)\right]. \tag{5}$$

Next we show that in a proper DSR, the expert's score can (aside from degenerate cases) only depend on the utility of the outcome obtained, and not on the outcome itself.

Lemma 3.4. *Let s be a proper DSR and $\omega_1, \omega_2 \in \Omega$ be two outcomes with $u(\omega_1) = u(\omega_2)$. Let $Q \in \Delta(H), P_\alpha \in \Delta(\Omega)^H$ be s.t. for all $e \in H$*

$$\min_{\omega \in \Omega} u(\omega) < \mathbb{E}_{O \sim P_\alpha}[u(O) \mid e] < \max_{\omega \in \Omega} u(\omega). \tag{6}$$

Further, let $\omega_1, \omega_2 \in \operatorname{supp}(\hat{P}_\alpha(\cdot \mid e))$ for some $e \in H$. Then

$$s(Q, P_\alpha, e, \omega_1) = s(Q, P_\alpha, e, \omega_2). \tag{7}$$

Because of Lemmas 3.3 and 3.4, we will therefore limit our attention in the following to scoring rules $s\colon \Delta(\Omega) \times \mathbb{R}^H \times H \times \mathbb{R} \to \mathbb{R}$ that take as input a vector of reported expected utilities $\hat{\boldsymbol{\mu}} \in \mathbb{R}^H$ (instead of the full distribution $\hat{P}_\alpha$) and the utility $y \in \mathbb{R}$ of the outcome obtained (instead of the outcome $\omega \in \Omega$ itself).

To simplify notation throughout the paper, we adopt a slightly unusual convention for reporting the conditional means. Specifically, we take an *honest* report of means to be

$$\hat{\mu}_e = \mu_e := \mathbb{E}_{E \sim Q, O \sim P_\alpha}[\mathbb{1}[E{=}e]\, u(O)] = Q(e)\mathbb{E}_{O \sim P_\alpha}[u(O) \mid e], \tag{8}$$

where we will call μ_e the true means. In words, we would like the expert to report for each e the expected utility if the principal only received utility in the case that $E = e$. Of course, since the expert reports both $\hat{Q}$ and $\hat{\boldsymbol{\mu}}$, we can translate a reported mean of $\hat{\mu}_e$ for e into the more traditional notion of conditional expected utility by dividing by $\hat{Q}(e)$, except for the degenerate case $\hat{Q}(e) = 0$.

4 Non-negative affinity

Lemma 4.1. *Let s be a proper DSR. Then there are functions $f_Y\colon \Delta(H) \times \mathbb{R}^H \to \mathbb{R}^H_{\geq 0}, f_E\colon \Delta(H) \times \mathbb{R}^H \to \mathbb{R}^H$, $g\colon \Delta(H) \times \mathbb{R}^H \to \mathbb{R}$, s.t. for all $\hat{Q}, \hat{\boldsymbol{\mu}}, e, y$,*

$$s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, y) = f_Y(\hat{Q}, \hat{\boldsymbol{\mu}})_e y + f_E(\hat{Q}, \hat{\boldsymbol{\mu}})_e + g(\hat{Q}, \hat{\boldsymbol{\mu}}).$$

We are essentially claiming that s is affine in the true $Q, \boldsymbol{\mu}$ – see below. Note that we could do away with the function g and incorporate it into f_E. However, below it will be useful to separate the two.

Corollary 4.2. *Let s be a proper scoring rule specified via f_Y, f_E, g as per Lemma 4.1. Then for all reports $\hat{Q}, \hat{\boldsymbol{\mu}}$ evidence variables E distributed according to Q and all means Y with true means $\boldsymbol{\mu}$,*

$$\mathbb{E}\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, E, Y)\right] = (f_E(\hat{Q}, \hat{\boldsymbol{\mu}}), f_Y(\hat{Q}, \hat{\boldsymbol{\mu}}))(Q, \boldsymbol{\mu}) + g(\hat{Q}, \hat{\boldsymbol{\mu}}).$$

Note that to obtain that the expectation of s is affine in $(Q, \boldsymbol{\mu})$, it is necessary that $\boldsymbol{\mu}$ follows our convention. Without the convention, we would get a non-linear term (the products of $Q(e)$ and the standard expected utility of Y given e). For the rest of the paper it is essential that we can work with linear terms.

Corollary 4.2 allows us to introduce some additional simplifying notation. For a fixed scoring rule s with functions f_Y, f_E, g we let $f = (f_Y, f_E)$ and $s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) := f(\hat{Q}, \hat{\boldsymbol{\mu}})(Q, \boldsymbol{\mu}) + g(\hat{Q}, \hat{\boldsymbol{\mu}})$, which in turn by Corollary 4.2 is the expected score under true means $\boldsymbol{\mu}$ and $E \sim Q$.

5 Characterizations

So far we have established that (up to edge cases) proper decision scoring rules are affine in the true evidence distribution Q and the true means $\boldsymbol{\mu}$. The question of characterizing affine scoring rules

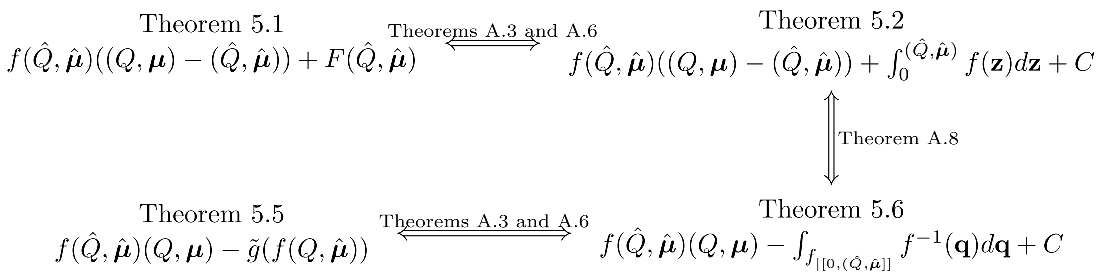

Figure 1: An overview of our characterizations of proper decision scoring rules and their relations

has been considered before in different settings. In particular, Frongillo and Kash (2014) study the question in its generic form. Also, proper scoring rules for prediction are trivially affine in the true distribution. Thus, characterizing proper scoring rules for prediction (as done by, e.g., Gneiting and Raftery, 2007) is another special case of characterizing proper affine scoring rules. We can therefore apply existing ideas to characterize proper decision scoring rules. We do this in Section 5.1, obtaining two different characterizations. In Section 5.2, we then give two other characterizations that we have not seen in this form before. We find one of them (Theorem 5.6) to be particularly intuitive. In Figure 1, we give an overview of our characterizations and how they are related via some of the generic convex analysis results of Appendix A.

5.1 Characterization à la Gneiting and Raftery (2007) and Frongillo and Kash (2014)

Theorem 5.1. *Let s be a DSR. Then s is proper if and only if there exist functions $f\colon \Delta(H_{-i})\times \mathbb{R}^{H_{-i}} \to \mathbb{R}^{H_{-i}} \times \mathbb{R}_{\geq 0}^{H_{-i}}$ and $F\colon \Delta(H_{-i})\times\mathbb{R}^{H_{-i}}\to\mathbb{R}$ s.t.*

$$s(\hat{Q},\hat{\boldsymbol{\mu}},Q,\boldsymbol{\mu}) = f(\hat{Q},\hat{\boldsymbol{\mu}})((Q,\boldsymbol{\mu})-(\hat{Q},\hat{\boldsymbol{\mu}}))+F(\hat{Q},\hat{\boldsymbol{\mu}}), \tag{9}$$

where F is convex and f is a subgradient of F.

This follows directly from our affinity results in Section 4 combined with the general characterization of linear scoring rules by Frongillo and Kash (2014). For completeness, we expand the relevant parts of the proof of their result in the appendix.

As noted by Gneiting and Raftery (2007) and Frongillo and Kash (2014), we can use Theorem A.3 to obtain the following alternative characterization.

Theorem 5.2. *Let s be a DSR. Then s is proper if and only if there is a cyclically monotone increasing function $f\colon \Delta(H)\times\mathbb{R}^H\to\mathbb{R}^H\times\mathbb{R}^H_{\geq 0}$, $C\in\mathbb{R}$, $\mathbf{b}\in\Delta(H)\times\mathbb{R}^H$ s.t.*

$$s(\hat{Q},\hat{\boldsymbol{\mu}},Q,\boldsymbol{\mu}) = f(\hat{Q},\hat{\boldsymbol{\mu}})((Q,\boldsymbol{\mu})-(\hat{Q},\hat{\boldsymbol{\mu}}))+\int_{\mathbf{b}}^{(\hat{Q},\hat{\boldsymbol{\mu}})} f(\mathbf{z})d\mathbf{z}+C. \tag{10}$$

The integral here is a path integral, as discussed in Appendix A.

Example 5. We obtain the simplest proper DSR that is strict w.r.t. evidence prediction and means by setting $f(\hat{Q},\hat{\boldsymbol{\mu}})=(\hat{Q},\hat{\boldsymbol{\mu}})$ (and $C=\int_0^{\mathbf{b}} f(\mathbf{z})d\mathbf{z}$), which yields

$$s(\hat{Q},\hat{\boldsymbol{\mu}},e,y)=\hat{Q}(e)+\hat{\mu}_e y-\frac{1}{2}\sum_{e'\in H}\hat{Q}(e')^2+\hat{\mu}_{e'}^2.$$

Note that the scoring of $\hat{Q}$ is exactly the Brier scoring rule of Example 2.

5.2 Offering different quantities of shares – the inverse of f as a pricing schedule

Lemma 5.3. *Let s be a proper DSR with $s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) = f(\hat{Q}, \hat{\boldsymbol{\mu}})\boldsymbol{\mu} - g(\hat{Q}, \hat{\boldsymbol{\mu}})$ for all $\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}$. Then for all $(\hat{Q}_1, \hat{\boldsymbol{\mu}}_1), (\hat{Q}_2, \hat{\boldsymbol{\mu}}_2)$, $f(\hat{Q}_1, \hat{\boldsymbol{\mu}}_1) = f(\hat{Q}_1, \hat{\boldsymbol{\mu}}_2) \implies g(\hat{Q}_1, \hat{\boldsymbol{\mu}}_1) = g(\hat{Q}_1, \hat{\boldsymbol{\mu}}_2)$.*

By Lemma 5.3, any DSR can be interpreted as a set of offers of getting $f(Q, \hat{\boldsymbol{\mu}})$ shares at a price of $g(Q, \hat{\boldsymbol{\mu}})$. By reporting a $(\hat{Q}, \hat{\boldsymbol{\mu}})$, the player accepts exactly one of these offers. Instead of having the expert report $\hat{\boldsymbol{\mu}}$, we will imagine that the expert chooses one of the offers by choosing a quantity $\mathbf{q}(= f(\hat{Q}, \hat{\boldsymbol{\mu}})) \in \mathrm{im}(f)$ of Arrow-Debreu securities in different evidence values e and shares in the principal projects that pay conditional on different e. (The set $\mathrm{im}(f) \subseteq \mathbb{R}^H \times \mathbb{R}^H_{\geq 0}$ is the set of possible quantity vectors available for sale.) Of course, there may be multiple true vectors of expected utilities $\boldsymbol{\mu}$ under which the expert prefers the same quantity $\mathbf{q}$. However, by Lemma 5.3, the quantity $\mathbf{q}$ of shares uniquely determines the price. (If there were multiple prices, the expert would always choose the lowest price.) We define $\tilde{g}(\mathbf{q}) = g(f^{-1}(\mathbf{q}))$ to be the (by Lemma 5.3 unique) price of $\mathbf{q}$ shares.

Lemma 5.4. *Let $\tilde{g}$ be the quantity-price function of a proper DSR. Then g is convex.*

Now for any $\mathbf{q}$, $f^{-1}(\mathbf{q}) \subseteq \Delta(H) \times \mathbb{R}^H$ is the set of true mean vectors $(Q, \boldsymbol{\mu})$ under which it is rational for the expert to select $\mathbf{q}$. Note that even if we had not defined f and had instead viewed our scoring rule as a list of offers of $\mathbf{q}$ shares for a price of $\tilde{g}(\mathbf{q})$, we could define f^{-1} as the function that infers beliefs from the choice of $\mathbf{q}$. It is left to show that f^{-1} must be a (set-valued) subgradient of the pricing function $\tilde{g}$. This is an easy task. Notice that for all $\mathbf{q}_0, \mathbf{q}$

$$\mathbf{q}_0 f^{-1}(\mathbf{q}_0) - \tilde{g}(\mathbf{q}_0) \geq \mathbf{q} f^{-1}(\mathbf{q}_0) - \tilde{g}(\mathbf{q}), \tag{11}$$

where in case $f^{-1}(\mathbf{q}_0)$ is multi-valued, the inequality should be understood as applying to each pair of an element of the left set and and an element of the right set. This exactly expresses the property of f^{-1}: Whenever the true $(Q, \boldsymbol{\mu})$ is from $f^{-1}(\mathbf{q}_0)$, the expert (weakly) prefers buying $\mathbf{q}_0$ (over any other quantity $\mathbf{q}$). Now rearranging Ineq. 11 gives $\tilde{g}(\mathbf{q}) \geq \tilde{g}(\mathbf{q}_0) + f^{-1}(\mathbf{q}_0)(\mathbf{q} - \mathbf{q}_0)$. With this, we get the following characterization.

Theorem 5.5. *Let s be a DSR. Then s is (strictly) proper if and only if there are $f, \tilde{g}$ s.t.*

$$s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) = f(\hat{Q}, \hat{\boldsymbol{\mu}})(Q, \boldsymbol{\mu}) - \tilde{g}(f(Q, \hat{\boldsymbol{\mu}})), \tag{12}$$

where $f \colon \mathbb{R}^H \to \mathbb{R}^H_{\geq 0}$, $\tilde{g} \colon \mathrm{im}(f) \to \mathbb{R}$ is (strictly) convex and f^{-1} is a subgradient of $\tilde{g}$.

Theorem 5.6. *Let s be a DSR. Then s is (strictly) proper if and only if there is cyclically monotone increasing $f \colon \Delta(H) \to \mathbb{R}^H \to \mathbb{R}^H \times \mathbb{R}^H_{\geq 0}$, $C, \mathbf{b}$ s.t.*

$$s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) = f(\hat{Q}, \hat{\boldsymbol{\mu}})(Q, \boldsymbol{\mu}) - \int_{f_{|[\mathbf{b}, (\hat{Q}, \hat{\boldsymbol{\mu}}]}} f^{-1}(\mathbf{q}) d\mathbf{q} + C. \tag{13}$$

Note that this is a path integral again. Since $f_{|[\mathbf{b}, (\hat{Q}, \hat{\boldsymbol{\mu}}]}$ might not be a continuous path and f^{-1} may not be single-valued, we need a minor extension of the notion of path integral, see Appendix A.5.

We now have a few different ways of proving this: Theorem 5.6 follows immediately from Theorem 5.5 and Theorem A.3 (which states that up to a constant a convex function is equal to the path integral over its subgradient). Theorem 5.6 also follows immediately from Theorem 5.2 and Theorem A.8 (the result on the integral of the inverse). Finally, for a direct proof we could modify our direct proof of Theorem 5.2. Roughly, instead of proving Lemma D.3, we could directly derive the integral-of-the-inverse shape, using parts of the proof of Lemma D.3 and the beginning of the proof of Theorem A.8 (cf. footnote 3).

6 Characterizations for special cases

6.1 Principal without private evidence ($|H| = 1$)

As noted in the discussion in Section 2, in most single-expert scenarios the principal would simply reveal all her private evidence to the expert before asking the expert for a prediction. Mathematically, this is the case where $|H| = 1$. We can here write DSRs as functions $s\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}\colon \hat{\mu}, y \mapsto s(\hat{\mu}, y)$. Since functions from $\mathbb{R} \to \mathbb{R}$ are cyclically monotone increasing if and only if they are monotone increasing, and because the path integral becomes the regular Riemann integral, one of our characterizations (Theorem 5.2) becomes especially simple in that it does not require any concepts that are beyond an introductory course in analysis:

Corollary 6.1. *A DSR $s(\hat{\mu}, y)$ for the case $|H| = 1$ is proper if and only if there is a monotonically non-decreasing function $f\colon \mathbb{R} \to \mathbb{R}_{\geq 0}$ and $C \in \mathbb{R}$ s.t.*

$$s(\hat{\mu}, y) = f(\hat{\mu})(y - \hat{\mu}) + \int_0^{\hat{\mu}} f(x)dx + C.$$

Moreover, s is strictly proper w.r.t. the recommendation if $f > 0$ and strictly proper w.r.t. the mean if and only if f is strictly monotonically increasing.

We also give the $|H| = 1$ version of Theorem 5.6:

Corollary 6.2. *A DSR $s(\hat{\mu}, y)$ for the case $|H| = 1$ is proper if and only if there is a monotonically non-decreasing function $f\colon \mathbb{R} \to \mathbb{R}_{\geq 0}$ and $C \in \mathbb{R}$ s.t.*

$$s(\hat{\mu}, y) = f(\hat{\mu})y - \int_{f(0)}^{f(\hat{\mu})} f^{-1}(z)dz + C.$$

Moreover, s is strictly proper w.r.t. the recommendation if $f > 0$ and strictly proper w.r.t. the mean if and only if f is strictly monotonically increasing.

In terms of techniques used, this looks as innocent as Corollary 6.1. Note, however, that we do require the extensions of Appendix A.5 for the integral to be well-defined if f is discontinuous or not strictly increasing.

7 Multiple experts

Besides eliciting from a single expert, we are interested in designing mechanisms for eliciting information for decision making from multiple experts, perhaps akin to real-world prediction markets. We are here interested in generic mechanisms, that is, mechanisms that work regardless of how the experts' information is structured. We will show by a revelation principle-type result that such mechanisms are characterized by proper decision scoring rules. Thus, our results (in particular, the characterizations from Section 5) characterize generic proper mechanism for eliciting information for decision making.

7.1 Expert information structure

Again, we consider a principal who selects from a set of actions A. After she has taken an action, an outcome from Ω is obtained. The principal would like to select the action that maximizes the expectation of the value of some utility function $u\colon \Omega \to \mathbb{R}$. This time the principal consults n different experts and (to keep things simple) does not have any private evidence of her own. Again, she asks for information, then takes the best action given the information submitted and finally rewards the experts based on the submitted information and the outcome obtained.

When it comes to the format and reporting of information, however, switching to the multi-expert setting poses a few additional challenges compared to the single-expert setting. First, of what types are the beliefs and reports of the experts? We do not want to simply let each expert's beliefs be some conditional probability distribution $\Delta(\Omega)^A$ again, because it would be unclear how one would aggregate these beliefs. Again, we make use of a standard solution from the economic literature: the common prior model.

We assume that each expert $i = 1, .., n$ has access to a private piece of information from some set H_i. The experts (and principal) share a common prior $Q \in \Delta(H)$ over $H := \bigtimes_i H_i$. For simplicity, we assume that $Q(\mathbf{e}) > 0$ for all $\mathbf{e} \in H$ so that the experts cannot contradict each other. We imagine also that the experts (and principal) share a common conditional distribution $P \in \Delta(\Omega)^{A \times H}$, which for any action $a \in A$ and evidence vector $\mathbf{e} \in H$ specifies a probability distribution $P(\cdot \mid a, \mathbf{e})$ over outcomes given that $\mathbf{e}$ is observed and action a is taken. As a report, each expert – after observing $e_i \in H_i$ – submits $\hat{e}_i \in H_i$ and the principal chooses an action that is best given the reported evidence, i.e., an action from $\text{opt}_P(\hat{\mathbf{e}}) := \arg\max_{a \in A} \mathbb{E}_P [u(O) \mid \hat{\mathbf{e}}, a]$.

7.2 Truthful mechanisms

In general, a mechanism for a given private information structure is a special type of game Γ of n players in which each player observes some $e_i \in H_i$, at some point an action $a \in A$ is selected, and an outcome $\omega \sim P(\cdot \mid e_1, ..., e_n, a)$ is observed. Each player i's payoff function can be arbitrarily determined by Γ. We will denote it by u_i (not to be confused with u, which we will continue to use to denote the principal's utility function).

We say that a mechanism is truthful for a given information structure if the game has a Nash equilibrium $\boldsymbol{\sigma}$ s.t. in $\boldsymbol{\sigma}$ an optimal action (i.e., an element from $\arg\max_{a \in A} \mathbb{E}_P [u(O) \mid \mathbf{e}, a]$, where $\mathbf{e} \in H$ is the true observed information) is selected. For strict propriety, we could add additional restrictions. For example, for strict propriety w.r.t. evidence prediction, we could require that there is an *interpretation* function that takes a trajectory of the game as an argument and in $\boldsymbol{\sigma}$ accurately returns each player i's distribution $Q_i(\cdot \mid e_i) \in \Delta(H_{-i})$; and further require that if any player i deviates from σ_i in a way that misleads the interpretation function, player i's payoff decreases strictly relative to $\boldsymbol{\sigma}$.

Proper DSRs as characterized in this paper can be used in various ways to construct generic truthful mechanisms. Here is one example of such a setup:

Example 6. Take proper DSRs $s_1, ..., s_n$, where $s_i \colon \Delta(H_{-i}) \times \mathbb{R}^{H_{-i}} \times H_{-i} \times \Omega \to \mathbb{R}$. As usual when considering such a scoring rule, each expert i is asked to submit a recommendation function $\alpha_i \colon H_{-i} \to A$, an evidence prediction $\hat{Q}_i$ and a collection of means $\hat{\boldsymbol{\mu}}_i \in \mathbb{R}^{H_{-i}}$. In addition, each expert submits e_i itself. The principal then feeds into each submitted α_i the vector $\mathbf{e}_{-i}$ of evidence values submitted by the other experts. Of the resulting n recommendation, she selects one, say, according to majority vote (with arbitrary tie-breaking), and an outcome with utility y is obtained. Each expert i then receives the score $s_i(\hat{Q}_i, \hat{\boldsymbol{\mu}}_i, \mathbf{e}_{-i}, y)$. This scoring system is proper for every information structure, i.e., honest reporting is always a Nash equilibrium of this game.

It is also easy to come up with truthful mechanisms that score quite differently, at least in some situations. Here is one such example:

Example 7. The information structure is as follows. There are two experts. With probability 1/2, Expert 1 observes what the best action is. Otherwise, he observes nothing. Expert 2 observes (with probability 1) for each action a the outcome distribution $P(\cdot \mid a) \in \Delta(\Omega)$. The principal asks both experts to report their private information. If Expert 1 submits a recommendation $\hat{a} \in A$, the principal always follows that recommendation to obtain an outcome ω. Expert 1 is paid in proportion to $u(\omega)$, and Expert 2 is paid using the Brier score for his outcome prediction for $\hat{a}$ (cf. Example 3). If, on the other hand, Expert 1 provides no recommendation, the principal follows

Expert 2's recommendation and rewards him using a proper DSR for the $|H| = 1$ case. This mechanism is truthful for the information structure outlined.

The example shows that if one expert reports a definitive claim that cannot be overruled by some other expert, the principal can potentially score the latter expert in ways that do not correspond to proper DSRs.

7.3 Proper DSRs characterize truthful mechanisms in generic situations

Definition 5. We say that an information structure is *generic for i under observation of $H'_{-i} \subseteq H_i$* if player i's private evidence can imply any distribution Q_i over H'_{-i} and any family of distributions $(P(\cdot \mid \mathbf{e}_{-i}, a))_{a,\mathbf{e}_{-i}}$.

Theorem 7.1. *Let Γ in NE $\boldsymbol{\sigma}$ be a proper mechanism for an information structure that is generic for i under observation of H'_{-i}. Then the following is a proper decision scoring rule:*

$$\Delta(H_{-i}) \times \Delta(\Omega)^{\Delta(H_{-i}) \times A} \times H_{-i} \times \Omega \to \mathbb{R} \colon (\hat{Q}_i, \hat{P}_i, e_{-i}, \omega) \mapsto u_i(\boldsymbol{\sigma}, (\mathbf{e}_{-i}, (\hat{Q}_i, \hat{P}_i)), \omega). \quad (14)$$

We now give some intuition for the core of the genericism condition above and for what the theorem says. Intuitively, in an elicitation process it can happen that players $-i$ honestly make some definitive claim that they are certain cannot be overruled by whatever player i reports. If this happens, then Theorem 7.1 does not apply and, depending on the exact nature of the definitive information, the mechanism might be able to reward i in specific ways that violate propriety in general, see Example 7. Our theorem applies when players $-i$ provide information that is tentative and leaves open (if only with small probability) that player i can hold any belief about what the best action is and what the outcome distribution over actions is.

Of course, situations like Example 7 may well occur – that is, in some cases at least *some* of the consulted experts can supply only a specific type of information and can therefore be scored in problem-specific ways. Also, the extreme assumption of e_{-i} leaving *everything* open is usually not realistic, either. Most of the time, experts may well be able to definitely rule out various absurd reports. Nevertheless, we find that Theorem 7.1 operates on a useful model.

7.4 Some conclusions about how to design realistic mechanisms

We can now draw conclusions from these results about what kind of characteristics any proper mechanism for eliciting information for decision making from multiple experts must have in generic situations. For instance, as in the single-expert case, it shows that we cannot incentivize experts to – along with an honest recommendation – reveal anything other than the expected utility of taking the recommended action. Further, no expert may profit from the failure of the principal's project. If we imagine the principal to be a firm maximizing its value, then no expert can be allowed to short-sell shares in the firm. In the rest of this section, we consider another, multi-expert-specific, desirable property that as a consequence of our results we cannot obtain.

We might like to reward experts in proportion to how much their report updates the principal's beliefs. This is one of many desirable properties of prediction markets: experts (or traders) are rewarded based on how far they can move the market probabilities toward the truth. For example, an expert who at any point simply agrees with the market probabilities (because he has no relevant private information) can earn no money (in expectation). An expert who updates the market probability for an event from, say, 0.5 to 0.1 receives a high score in expectation (assuming 0.1 represents his true beliefs over the outcome of the random event). Rewarding experts for their impact on the market probability has many advantages. For instance, it sets an incentive to acquire relevant information. Therefore, we might want an elicitation mechanism for decision making – perhaps a kind of decision market (see Section 8.3) – to similarly reward experts for submitting evidence that yields large (justified) changes in the principal's beliefs.

However, from our results it follows immediately that a number of types of changes cannot be rewarded at all. An expert's score cannot depend on how much the trader's report moves the distributions for suboptimal actions. Experts also cannot be rewarded for changing what action is recommended. Generally, if two pieces of information $e_i^1, e_i^2 \in H_i$ have the same implications for the expected utility given the best action (and make the same predictions about what evidence the other experts submit), the expert receives the same expected score from honestly reporting e_i^1 and honestly reporting e_i^2. This is the case even if e_i^1 affects what the best action is and implies wild changes to the distributions of all actions while e_i^2 does not change the principal's beliefs at all. In particular this implies that a generic strictly proper DSR gives positive expected rewards even to experts i whose private evidence E_i turns out to be of no value to the principal.

How can the principal make sure that despite these impossibilities, experts with more useful information receive higher scores? The only way out, it seems, is to reward experts based on the *ex ante* value of their information. That is, pay expert i (in shares or constant reward) in proportion to how much the principal would be willing to pay to learn E_i. One could also use the willingness to pay given that one already knows or will know $\mathbf{E}_{-i}$. In the extreme case, one could even give a constant score of 0 to experts whose value of information is zero. (The mechanism would then not quite satisfy our generic notion of strict propriety anymore.) This way, obtaining E_i is incentivized to the extent that E_i is useful to the expert.

8 Related work

8.1 Othman and Sandholm (2010)

As far as we can tell, Othman and Sandholm (2010) are the first to consider designing proper decision scoring rules as considered in our paper. They study a simplified case in which $|\Omega| = 2$ and $|H| = 1$. Note that the two-outcome-case is special because the mean of a binary random variable fully determines its distribution. In Section 2.3.2, they give a characterization of *differentiable* proper decision scoring rules. A generalization to differentiable scoring rules $s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu})$ is given in Appendix E.

8.2 Chen, Kash, Ruberry, et al. (2014)

Chen, Kash, Ruberry, et al. (2014) also characterize scoring rules for decision making (an alternative proof of this characterization is also given by Frongillo and Kash, 2014, Section E.1). Their setting is more general than ours in that they allow arbitrary *decision rules*. That is, they allow principals who do not choose the best action, but, for instance, randomize over the best few actions with probabilities depending on the expert's prediction. Randomizing over *all* actions in particular, even if not uniformly, allows any proper scoring rule for mere prediction to be used to construct strictly proper scoring rules for decision making (Chen, Kash, Ruberry, et al., 2014, Section 4).

They also characterize a much larger class of scoring rules: they merely require that the expert honestly reports the probability distribution that the recommendation gives rise to and allow scoring rules which strictly incentivize misreporting what the best action is. For principals who (like those in our setting) deterministically choose the best action according to the expert's report, their result is especially easy to derive and understand. To elicit an honest report about the probability distribution resulting from taking the expert's recommended action, the principal can use any (strictly) proper scoring rule for mere prediction (as defined and characterized by, e.g., Gneiting and Raftery, 2007, Section 2). As an example, consider the quadratic (Brier) scoring rule (Example 3).

Chen, Kash, Ruberry, et al. (2014, Section 5) do also consider the goal of characterizing preferences over lotteries for which making the best recommendation can be incentivized. But they do not give a characterization of proper decision scoring rules for expected utility-maximizing principals or of what information can be extracted along with the best action.[2]

8.3 Decision markets

As far as we are aware, most work on eliciting decision-relevant information from multiple agents has focused on designing prediction-market-like mechanisms or "decision markets" (as opposed to considering the class of direct-revelation mechanisms discussed in this paper) (e.g. Berg and Rietz, 2003; Hanson, 1999, 2002, 2006, 2013, Section IV). Othman and Sandholm (2010, Section 3) are the first to point out incentive problems with this model. Our impossibility results can be seen as an extension of their result (though we have limited attention to mechanisms which guarantee full information aggregation, which may not a be a primary goal for the design of decision markets). Inspired by Othman and Sandholm's proof that decision markets sometimes set poor incentives, Teschner, Rothschild, and Gimpel (2017) conduct an empirical study in which human subjects took the roles of the experts (or "traders") to show that strategic reporting may not be a problem in decision markets in practice. Chen et al. (2011; 2014) show that by randomizing over all actions (potentially with a strong bias toward the optimal action) decision markets can, in some sense, be made to be analogous to prediction markets.

8.4 Direct elicitation of properties

Typically, when designing scoring rules for prediction (without the recommendation component) the goal is to elicit entire probability distributions over outcomes. But a recent line of work has explored the direct elicitation of particular properties of the distribution *without* eliciting the entire distribution (e.g. Abernethy and Frongillo, 2012; Bellini and Bignozzi, 2015; Gneiting, 2011; Lambert, Pennock, and Shoham, 2008). Of course, in principle, one could elicit entire distributions and would thereby elicit all properties. But eliciting, say, a single-valued point forecast may be required "for reasons of decision making, market mechanisms, reporting requirements, communications, or tradition, among others" (Gneiting, 2011, Section 1). Lemma 3.3 gives another reason to study scoring rules for eliciting just the expected utility, albeit with the additional requirements that the expected score under honest reporting must be the same for two variables with equal mean (Lemma 3.2) and that the expected score under honest reporting must be increasing in the true mean of the random variable. Results from the literature on property elicitation can also be used to replace parts of the proof of our main result.

References

Abernethy, Jacob D. and Rafael M. Frongillo (June 2012). "A Characterization of Scoring Rules for Linear Properties." In: *Proceedings of the 25th Annual Conference on Learning Theory*. Ed. by Shie Mannor, Nathan Srebro, and Robert C. Williamson. Vol. 23. Proceedings of Machine Learning Research. Edinburgh, Scotland: PMLR, pp. 27.1–27.13. URL: `http://proceedings.mlr.press/v23/abernethy12.html`.

Agrawal, Shipra et al. (2009). "A unified framework for dynamic pari-mutuel information market design." In: *EC '09 Proceedings of the 10th ACM conference on Electronic commerce*, pp. 255–264.

Bellini, Fabio and Valeria Bignozzi (2015). "On elicitable risk measures." In: *Quantitative Finance* 15.5, pp. 725–733. DOI: `10.1080/14697688.2014.946955`.

[2] In earlier, work they mention in passing (Chen and Kash, 2011, Section 6) that the scoring rule of Example 5 strictly incentivizes the expert to not only recommend the best action but also the distribution resulting from taking that action. But this is not the case, as the scoring rule depends only on the reported mean. More generally, we have shown in Lemma 3.3 that no right-action proper scoring rule can strictly incentivize honesty about anything other than the mean of the recommended action.

Berg, Joyce E. and Thomas A. Rietz (2003). "Prediction Markets as Decision Support Systems." In: *Information Systems Frontiers* 5.1, pp. 79–93. DOI: 10.1023/A:1022002107255.

Brier, Glenn W. (Jan. 1950). "Verification of Forecasts Expressed in Terms of Probability." In: *Monthly Weather Review* 78.1.

Chen, Yiling and Ian A. Kash (2011). "Information elicitation for decision making." In: *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011), Taipei, 2-6 May 2011, edited by Tumer, Yolum, Sonenberg, and Stone.* Springer. URL: https://dash.harvard.edu/bitstream/handle/1/5128480/chen_aamas11_0.pdf?sequence=1.

Chen, Yiling, Ian A. Kash, Mike Ruberry, et al. (2011). "Decision markets with Good Incentives." In: *Lecture Notes in Computer Science 7090*, pp. 72–83. URL: https://dash.harvard.edu/bitstream/handle/1/11693950/Chen_DecisionMarkets.pdf?sequence=1.

— (June 2014). "Eliciting Predictions and Recommendations for Decision Making." In: *ACM Transactions on Economics and Computation.* Vol. 2. 2. Chap. 6. URL: http://yiling.seas.harvard.edu/wp-content/uploads/a6-chen.pdf.

Chen, Yiling and David M. Pennock (2007). "A utility framework for bounded-loss market makers." In: *UAI'07 Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence*, pp. 49–56.

Chen, Yiling and Jennifer Wortman Vaughan (2010). "A new understanding of prediction markets via no-regret learning." In: *EC' 10 Proceedings of the 11th ACM conference on Electronic commerce*, pp. 189–198.

Frongillo, Rafael M. and Ian A. Kash (2014). *General Truthfulness Characterizations Via Convex Analysis.* Short version published in WINE 2014. URL: https://www.cs.colorado.edu/~raf/media/papers/wine-affine.pdf.

Gneiting, Tilmann (2011). "Making and Evaluating Point Forecasts." In: *Journal of the American Statistical Association* 106.494, pp. 746–762. DOI: 10.1198/jasa.2011.r10138.

Gneiting, Tilmann and Adrian E. Raftery (Mar. 2007). "Strictly Proper Scoring Rules, Prediction, and Estimation." In: *Journal of the American Statistical Association* 102.477. DOI: 10.1198/016214506000001437.

Good, I. J. (1952). "Rational Decisions." In: *Journal of the Royal Statistical Society. Series B (Methodological)* 14 (1), pp. 107–114. URL: https://www.jstor.org/stable/2984087.

Hanson, Robin (1999). "Decision Markets." In: *IEEE Intelligent Systems* 14.3, pp. 16–19. URL: http://mason.gmu.edu/~rhanson/decisionmarkets.pdf.

— (2002). "Decision Markets." In: *Entrepreneurial Economics – Bright Ideas from the Dismal Science.* Ed. by Alexander Tabarrok. Oxford University Press. Chap. 5, pp. 79–85.

— (2003). "Combinatorial Information Market Design." In: *Information Systems Frontiers* 5.1, pp. 107–119.

— (Nov. 2006). "Decision Markets for Policy Advice." In: *Promoting the General Welfare: New Perspectives on Government Performance.* Ed. by Alan S. Gerber and Eric M. Patashnik. Brookings Institution Press, pp. 151–173.

— (2007). "Logaritmic Market Scoring Rules for Modular Combinatorial Information Aggregation." In: *The Journal of Prediction Markets* 1, pp. 3–15.

— (2013). "Shall We Vote on Values, But Bet on Beliefs?" In: *Journal of Political Philosophy* 21.2. DOI: 10.1111/jopp.12008.

Hartline, Jason D. et al. (July 2020). *Optimization of Scoring Rules.* URL: https://arxiv.org/pdf/2007.02905.pdf.

Key, Eric (Mar. 1994). "Disks, Shells, and Integrals of Inverse Functions." In: *The College Mathematics Journal* 25.2, pp. 136–138.

Lambert, Nicolas, David M. Pennock, and Yoav Shoham (2008). "Eliciting Properties of Probability Distributions." In: *ACM Conference on Electronic Commerce.* Extended Abstract.

McCarthy, John (Sept. 1956). “Measures of the Value of Information.” In: *Proceedings of the National Academy of Sciences of the United States of America* 42 (9), pp. 654–655. DOI: 10.1073/pnas.42.9.654. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC534271/.
Neyman, Eric, Georgy Noarov, and S. Matthew Weinberg (Feb. 2020). *Binary Scoring Rules that Incentivize Precision.* URL: https://arxiv.org/pdf/2002.10669.pdf.
Osband, Kent (Oct. 1989). “Optimal Forecasting Incentives.” In: *Journal of Political Economy* 97.5, pp. 1091–1112. URL: https://www.jstor.org/stable/1831887.
Othman, Abraham and Tuomas Sandholm (2010). “Decision Rules and Decision Markets.” In: *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010), van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada*, pp. 625–632.
Pennock, David M. and Rahul Sami (2007). “Computational Aspects of Prediction Markets.” In: *Algorithmic Game Theory.* Ed. by Noam Nisan et al. Cambridge University Press. Chap. 26, pp. 651–675.
Rockafellar, R. Tyrrell (1970). *Convex Analysis.* Princeton, NJ: Princeton University Press.
Savage, Leonard J. (Dec. 1971). “Elicitation of Personal Probabilities and Expectations.” In: *Journal of the American Statistical Association* 66 (336), pp. 783–801.
Teschner, Florian, David Rothschild, and Henner Gimpel (Sept. 2017). “Manipulation in Conditional Decision Markets.” In: *Group Decision and Negotiation* 26.5, pp. 953–971. DOI: 10.1007/s10726-017-9531-0.

A Background: convex analysis

All of our characterizations in Section 5 refer to either the concepts of convex functions and their subgradients; or the concepts of cyclic monotonicity and path integrals. We introduce these concepts and their relation here. Except for the ideas in Appendix A.5 (which is still based on well-known existing results), all of the below is known.

A.1 Convex functions and subgradients

A set $M \subseteq \mathbb{R}^n$ is convex if for all $\mathbf{x}, \mathbf{y} \in M$ and $p \in [0,1]$, $t\mathbf{x} + (1-t)\mathbf{y} \in \mathbb{M}$. In particular, $\mathbb{R}^n$ is convex for all $n \in \mathbb{N}$ and the set of probability distributions over any finite outcome space is convex. Further, the Cartesian product of two convex sets is convex.

Definition 6 (Convex function)**.** A function $F\colon M \to \mathbb{R}$ on a convex set M is *convex* if for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $p \in (0,1)$,

$$F(p\mathbf{x} + (1-p)\mathbf{y}) \leq pF(\mathbf{x}) + (1-p)F(\mathbf{y}).$$

We call F *strictly* convex if the inequality is strict for all $\mathbf{x}, \mathbf{y}, p$.

Definition 7 (Subgradient)**.** Let $F\colon \mathbb{R}^n \to \mathbb{R}$ be a function. We call $f\colon \mathbb{R}^n \to \mathbb{R}^n$ a *subgradient function* of F if for all $\mathbf{x}_0, \mathbf{x} \in \mathbb{R}^n$,

$$F(\mathbf{x}) \geq F(\mathbf{x}_0) + f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0). \tag{15}$$

Lemma A.1. *If a function F has a subgradient, F is convex.*

Proof. Let $\mathbf{z} \coloneqq p\mathbf{x} + (1-p)\mathbf{y}$ and f be F's subgradient. Then

$$\begin{aligned} F(p\mathbf{x} + (1-p)\mathbf{y}) &= F(\mathbf{z}) + f(\mathbf{z})(p\mathbf{x} + (1-p)\mathbf{y} - \mathbf{z}) \\ &= p(F(\mathbf{z}) + f(\mathbf{z})(\mathbf{x} - \mathbf{z})) + (1-p)(F(\mathbf{z}) + f(\mathbf{z})(\mathbf{y} - \mathbf{z})) \\ &\underset{\text{Ineq. 15}}{\leq} pF(x) + (1-p)F(y). \end{aligned}$$

□

A.2 Convex functions are path integrals of their subgradients

Theorem A.2 (Rockafellar, 1970, Corollary 24.2.1). *Let $\zeta\colon \mathbb{R} \to \mathbb{R}$ be convex with subderivative $\zeta'\colon \mathbb{R} \to \mathbb{R}$. Then for all $x, y \in \mathbb{R}$,*

$$\zeta(y) - \zeta(x) = \int_x^y \zeta'(t)dt. \tag{16}$$

This can be seen as a version of the fundamental theorem of calculus. Of course, the latter theorem is usually used to calculate a definite integral by finding the antiderivative of the function the integral is over. We will use it in the other direction, however, to replace a function with an integral over its subderivative.

Define $[\mathbf{x}_1, \mathbf{x}_1] \coloneqq \{\mathbf{x}_1 + t(\mathbf{x}_2 - \mathbf{x}_1) \mid t \in [0, 1]\}$ to be the line from $\mathbf{x}_1$ to $\mathbf{x}_2$.

Theorem A.3. *Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be the subgradient of some function $F\colon \mathbb{R}^n \to \mathbb{R}$. Then for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}$.,*

$$\int_{\mathbf{x}_1}^{\mathbf{x}_2} f(\mathbf{z})d\mathbf{z} = F(\mathbf{x}_2) - F(\mathbf{x}_1), \tag{17}$$

where the path integral is path-independent.

Similarly, this result is analogous to the fundamental theorem of calculus for line integrals (a.k.a. the gradient theorem). While Rockafellar (1970) only gives the single-dimensional version of the result, Theorem A.3 is easy to derive from the single-dimensional version as has been noted before (e.g. Frongillo and Kash, 2014, Appendix A, Fact 3).

Proof. For any $\mathbf{x}_1, \mathbf{y}_2 \in \mathbb{R}^n$ we can consider the restriction of f to . Using Lemma A.2, we get

$$\int_{L_{\mathbf{x}_1,\mathbf{x}_2}} f(\mathbf{z})d\mathbf{z} + \int_{L_{\mathbf{x}_2,\mathbf{x}_3}} f(\mathbf{z})d\mathbf{z} = F(\mathbf{x}_2) - F(\mathbf{x}_1) + F(\mathbf{x}_3) - F(\mathbf{x}_2) = F(\mathbf{x}_3) - F(\mathbf{x}_1) = \int_{L_{\mathbf{x}_1,\mathbf{x}_3}} f(\mathbf{z})d\mathbf{z}$$

as claimed. □

A.3 Cyclic monotonicity

Definition 8. A function $f\colon M \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is called *cyclically monotone increasing* if for all sequences $\mathbf{x}_0, ..., \mathbf{x}_n \in M$ with $n \geq 1$,

$$f(\mathbf{x}_0)(\mathbf{x}_1 - \mathbf{x}_0) + f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) + ... + f(\mathbf{x}_{n-1})(\mathbf{x}_n - \mathbf{x}_{n-1}) + f(\mathbf{x}_n)(\mathbf{x}_0 - \mathbf{x}_n) \leq 0. \tag{18}$$

We call f is *strictly* cyclically monotone increasing if this inequality is strict unless $\mathbf{x}_0 = ... = \mathbf{x}_n$.

Lemma A.4. *A function $f\colon M \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is cyclically monotone increasing if and only if for all $\mathbf{z}_0, ..., \mathbf{z}_n \in M$,*

$$f(\mathbf{z}_1)(\mathbf{z}_1 - \mathbf{z}_0) + f(\mathbf{z}_2)(\mathbf{z}_2 - \mathbf{z}_1) + ... + f(\mathbf{z}_n)(\mathbf{z}_n - \mathbf{z}_{n-1}) + f(\mathbf{z}_0)(\mathbf{z}_0 - \mathbf{z}_n) \geq 0. \tag{19}$$

Proof. Multiplying Ineq. 18 by -1 yields

$$f(\mathbf{x}_0)(\mathbf{x}_0 - \mathbf{x}_1) + f(\mathbf{x}_1)(\mathbf{x}_1 - \mathbf{x}_2) + ... + f(\mathbf{x}_{n-1})(\mathbf{x}_{n-1} - \mathbf{x}_n) + f(\mathbf{x}_n)(\mathbf{x}_n - \mathbf{x}_0) \geq 0. \tag{20}$$

Choosing $\mathbf{x}_n = \mathbf{z}_0,\ \mathbf{x}_{n-1} = \mathbf{z}_1,\ ...,\ \mathbf{x}_0 = \mathbf{z}_n$, we obtain

$$f(\mathbf{z}_n)(\mathbf{z}_n - \mathbf{z}_{n-1}) + f(\mathbf{z}_{n-1})(\mathbf{z}_{n-1} - \mathbf{z}_{n-2}) + ... + f(\mathbf{z}_1)(\mathbf{z}_1 - \mathbf{z}_0) + f(\mathbf{z}_0)(\mathbf{z}_0 - \mathbf{z}_n) \geq 0. \tag{21}$$

And this is the same as Ineq. 19, except that the order of the first n summands is reversed. □

Lemma A.5. *Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be cyclically monotone increasing. Then f is path-independently integrable.*

Proof. First, note that for any $\boldsymbol{\mu}_a, \boldsymbol{\mu}_b$, each of the n entries of n are monotone on the line $L_{\boldsymbol{\mu}_a,\boldsymbol{\mu}_b}$. It follows that the components of f are integrable on $L_{\boldsymbol{\mu}_a,\boldsymbol{\mu}_b}$. (E.g., Rudin 1976, Theorem 6.9.) Hence, f is (path-)integrable on $L_{\boldsymbol{\mu}_a,\boldsymbol{\mu}_b}$.

It is left to show path-independence, i.e., that for closed curves γ,

$$\int_\gamma f(\mathbf{x}) d\mathbf{x} = 0.$$

This can be seen as follows. Because f, is cyclically monotone increasing, it is for all $\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_k \in \mathbb{R}^n$,

$$\sum_{i=0}^{k} f(\boldsymbol{\mu}_i)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i-1}) \geq 0 \geq \sum_{i=0}^{k} f(\boldsymbol{\mu}_{i-1})(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i-1}). \tag{22}$$

Now, because f is integrable, if we let the cycle $\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_k$ become arbitrarily fine approximations of γ, the left and right sum converge to the integral along γ and therefore to the same value. By Ineq. 22, that value must be 0. □

A.4 A function is a subgradient if and only if it is cyclically monotone increasing

Theorem A.6 (Rockafellar, 1970, Theorem 24.8)**.** *Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$. Then there is a function F such that f is a subgradient function of F if and only if f is cyclically monotone increasing.*

A.5 Path integral of the inverse

Besides taking integrals along lines $[\mathbf{a}, \mathbf{b}] \subseteq \mathbb{R}^n$, we would like to take curve integrals along functions $\gamma\colon [\mathbf{a}, \mathbf{b}] \to \mathbb{R}^n$. In particular, we want to take them along functions γ that are cyclically monotone increasing, but not necessarily continuous.

To make this well-defined we first extend γ to a set-valued function $\bar{\gamma}$ such that $\bar{\gamma}([\mathbf{a}, \mathbf{b}])$ is a curve. So let γ be discontinuous at $\mathbf{x} \in [\mathbf{a}, \mathbf{b}]$. Specifically imagine that there is a $\delta > 0$ s.t. for all $\epsilon > 0$,

$$\|\gamma(\mathbf{x} + \epsilon(\mathbf{b} - \mathbf{a})) - \gamma(\mathbf{x})\| > \delta,$$

i.e., that γ jumps immediately after $\mathbf{x}$. Imagine further that γ is continuous in the other direction from $\mathbf{x}$. Then define

$$\bar{\gamma}(\mathbf{x}) \coloneqq [\gamma(\mathbf{x}), \lim_{\epsilon \downarrow 0} \gamma(\mathbf{x} + \epsilon(\mathbf{b} - \mathbf{a}))]$$

If γ jumps to the left or on both sides of $\mathbf{x}$, we define $\bar{\gamma}(\mathbf{x})$ analogously. If γ is continuous at $\mathbf{x}$, we simply let $\bar{\gamma}(\mathbf{x}) = \{\gamma(\mathbf{x})\}$.

We can now define the path integral in almost the usual way via partitioning the curve $\bar{\gamma}$. So for each n let $\mathbf{y}_{n,0}, ..., \mathbf{y}_{n,n} \in \bar{f}([\mathbf{a}, \mathbf{b}])$ that are ordered in the natural way. Further let $\mathbf{y}_{n,0}, ..., \mathbf{y}_{n,n}$ become arbitrary fine as $n \to \infty$. We then consider limits

$$\sum_{i=1}^{n} g(\mathbf{y}_{n,i})(\mathbf{y}_{n,i} - \mathbf{y}_{n,i-1})$$

as $n \to \infty$. If these limits exist and are the same for all all partitions, we call that limit

$$\int_\gamma g(\mathbf{x}) d\mathbf{x}.$$

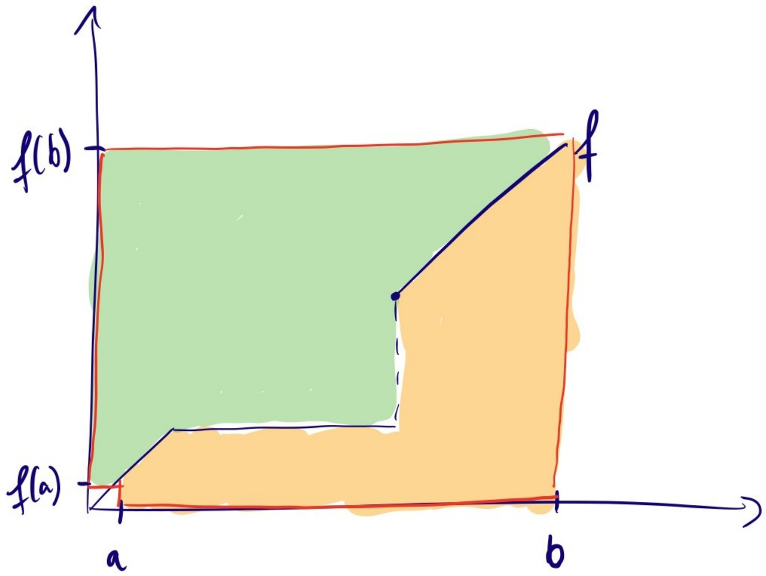

Figure 2: An illustration of Theorem A.8 for $n = 1$.

We have now slightly extended the notion of curves for curve integrals. Next, we would like to take integrals of the form

$$\int_{f_{|[\mathbf{a},\mathbf{b}]}} f^{-1}_{|[\mathbf{a},\mathbf{b}]}(\mathbf{y})d\mathbf{y},$$

where $f_{|[\mathbf{a},\mathbf{b}]}$ is the restriction of f to the line $[\mathbf{a},\mathbf{b}]$ and $f^{-1}_{|[\mathbf{a},\mathbf{b}]} \colon f([\mathbf{a},\mathbf{b}])$ is its inverse. This presents another small technical difficulty, which is that $f_{|[\mathbf{a},\mathbf{b}]}$ need not be injective and thus $f^{-1}_{|[\mathbf{a},\mathbf{b}]}$ may be set-valued. We will deal with this by tie-breaking to get a single-value. We will see that in our context it does not matter which value is chosen.

Lemma A.7. *Let f be cyclically monotone increasing. Then for all $\mathbf{y}, \mathbf{a}, \mathbf{b}$, $f^{-1}_{|[\mathbf{a},\mathbf{b}]}(\mathbf{y})$ is a line segment/interval.*

Theorem A.8. *Let $f \colon \mathbb{R}^n \to \mathbb{R}^n$ be cyclically monotone increasing and $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$. Then*

$$\int_{f_{|[\mathbf{a},\mathbf{b}]}} f^{-1}_{|[\mathbf{a},\mathbf{b}]}(\mathbf{y})d\mathbf{y} = \mathbf{b}f(\mathbf{b}) - \mathbf{a}f(\mathbf{a}) - \int_{\mathbf{a}}^{\mathbf{b}} f(\mathbf{x})d\mathbf{x}. \tag{23}$$

For strictly monotone, continuous functions $f \colon \mathbb{R} \to \mathbb{R}$ and regular Riemann integrals, this is a well-known and intuitive result (see, e.g., Key, 1994, Theorem 1). We generalize this result in two (novel, as far as we know) ways. The first is that we allow f to be only *weakly* monotone and discontinuous. As long as we keep $n = 1$, the result remains intuitive and we give the typical type of illustration in Figure 2. The integral over f from $\mathbf{a}$ to $\mathbf{b}$ is here just the yellow area under the curve. The integral over f^{-1} is simply the green area under the curve of f^{-1} from $f(\mathbf{a})$ to $f(\mathbf{b})$. Together, the two curves must equal the area demarcated by the red line, which is equal to $\mathbf{b}f(\mathbf{b}) - \mathbf{a}f(\mathbf{a})$. We further generalize the theorem by allowing multi-dimensional path integrals. Unfortunately, it is much harder to provide analogous pictures for the higher-dimensional case.

Proof. Consider any sequence of partitions $\mathbf{y}_{n,0}, ..., \mathbf{y}_{n,n}$ of the path $\bar{f}_{|[\mathbf{a},\mathbf{b}]}$ as specified above. First, take a sequence of partitions $\mathbf{x}_{n,0}, ..., \mathbf{x}_{n,n}$ of $[\mathbf{a},\mathbf{b}]$ (ordered in the natural way), s.t., for each $n \in \mathbb{N}, i \in \{0, ..., n\}$, $\mathbf{y}_{n,i} \in \bar{f}(\mathbf{x}_{n,i})$.

First, it is

$$\sum_{i=1}^{n} f^{-1}(\mathbf{y}_{n,i})(\mathbf{y}_{n,i}-\mathbf{y}_{n,i-1}) - \sum_{i=1}^{n} \mathbf{x}_{n,i}(\mathbf{y}_{n,i}-\mathbf{y}_{n,i-1}) \to 0 \text{ as } n \to \infty.$$

Note that the left-hand side is the "Riemann sum" for f^{-1} on the path f. Intuitively, this just means that whenever $f^{-1}(\mathbf{y}_{n,i})$ has multiple values, it doesn't matter which one we pick and so we can assume that we pick a specific $\mathbf{x}_i$ as per the above partition of $[\mathbf{a}, \mathbf{b}]$. This fact follows from Lemma A.7.

Furthermore, it is

$$\sum_{i=1}^{n} \mathbf{x}_{n,i}(\mathbf{y}_{n,i}-\mathbf{y}_{n,i-1}) - \sum_{i=1}^{n} \mathbf{x}_{n,i}(f(\mathbf{x}_{n,i}) - f(\mathbf{x}_{n,i-1})) \to 0 \text{ as } n \to \infty.$$

Now, we can rewrite this as

$$\begin{aligned}
& \sum_{i=1}^{n} \mathbf{x}_{n,i}(f(\mathbf{x}_{n,i}) - f(\mathbf{x}_{n,i-1})) \\
= \quad & \sum_{i=1}^{n} \mathbf{x}_{n,i} f(\mathbf{x}_{n,i}) - \mathbf{x}_{n,i-1} f(\mathbf{x}_{n,i-1}) - \sum_{i=1}^{n} (\mathbf{x}_{n,i} - \mathbf{x}_{n,i-1}) f(\mathbf{x}_{n,i-1}) \\
= \quad & \mathbf{x}_{n,n} f(\mathbf{x}_{n,n}) - \mathbf{x}_{0,0} f(\mathbf{x}_{0,0}) - \sum_{i=1}^{n} (\mathbf{x}_{n,i} - \mathbf{x}_{n,i-1}) f(\mathbf{x}_{n,i-1}) \\
\underset{n\to\infty}{\to} \quad & \mathbf{b} f(\mathbf{b}) - \mathbf{a} f(\mathbf{a}) - \int_{\mathbf{a}}^{\mathbf{b}} f(\mathbf{x}) d\mathbf{x},
\end{aligned}$$

as claimed. □

B Proofs for Section 3

B.1 Proof of Lemma 3.1

Lemma 3.1. *If s is proper, then $\mathbb{E}_{E\sim Q, O\sim P_\alpha}[s(Q,P,E,O)]$ is continuous in P, Q in the set of P, Q with full support.*

Proof. Because s is proper,

$$\mathbb{E}_{E\sim Q, O\sim P_\alpha}[s(Q,P,E,O)] = \max_{\hat{Q}\in\Delta(H), \hat{P}_\alpha \in \Delta(\Omega)^H} \mathbb{E}_{E\sim Q, O\sim P_\alpha}\left[s(\hat{Q}, \hat{P}_\alpha, E, O)\right].$$

For fixed, $\hat{Q}, \hat{P}_\alpha$, the term

$$\mathbb{E}_{E\sim Q, O\sim P_\alpha}\left[s(\hat{Q}, \hat{P}_\alpha, E, O)\right]$$

is affine in Q, P_α. Hence, $\mathbb{E}_{E\sim Q, O\sim P_\alpha}[s(Q,P,E,O)]$ is the point-wise maximum of a set of affine functions. It can be shown that the point-wise maximum of affine functions is convex. Finally, a convex function defined on an open interval is continuous on that interval. □

B.2 Proof of Lemma 3.2

Lemma 3.2. *Let s be a proper DSR and $P_\alpha, P'_{\alpha'} \in \Delta(\Omega)^H$ be s.t. for all $e \in H$*

$$\min_{\omega\in\Omega} u(\omega) < \mathbb{E}_{O\sim P_\alpha}[u(O) \mid e] = \mathbb{E}_{O\sim P'_{\alpha'}}[u(O) \mid e] < \max_{\omega\in\Omega} u(\omega). \quad (2)$$

Then

$$\mathbb{E}_{E\sim Q,O\sim P_\alpha}\left[s(Q,P_\alpha,E,O)\right]=\mathbb{E}_{E\sim Q,O\sim P_{\alpha'}}\left[s(Q,P_{\alpha'},E,O)\right]. \tag{3}$$

Proof. Let $\omega_L = \arg\min_{\omega\in\Omega} u(\omega)$ and $\omega_H = \arg\max_{\omega\in\Omega} u(\omega)$ (with ties broken arbitrarily). For any $\mathbf{p}\in(0,1)^H$, define $R_\mathbf{p}\in\Delta(\Omega)^H$ to be the distribution where for all $e\in H$:

$$R_\mathbf{p}(\omega_H \mid e) = p_e \tag{24}$$

$$R_\mathbf{p}(\omega_L \mid e) = 1-p_e \tag{25}$$

$$R_\mathbf{p}(\omega \mid e) = 0 \text{ for all } \omega\in\{\omega_L,\omega_H\} \tag{26}$$

Now consider any (non-extreme) P_α as well as $R_\mathbf{p}$ as defined above. For s to be proper it has to be

$$\mathbb{E}_{E\sim Q,O\sim P_\alpha}\left[s(Q,P_\alpha,E,O)\right]\geq\mathbb{E}_{E\sim Q,O\sim R_\mathbf{p}}\left[s(Q,R_\mathbf{p},E,O)\right] \tag{27}$$

whenever all the means of $R_\mathbf{p}$ are element-wise at most the means of P_α. The reverse inequality has to hold if the means of $R_\mathbf{p}$ are at at least as high the means of P_α. By continuity of $\mathbb{E}_{E\sim Q,O\sim R_\mathbf{p}}\left[s(R_{Q,\mathbf{p}},E,O)\right]$ as per Lemma 3.1, it follows that

$$\mathbb{E}_{E\sim Q,O\sim P_\alpha}\left[s(Q,P_\alpha,E,O)\right]=\mathbb{E}_{E\sim Q,O\sim R_\mathbf{p}}\left[s(Q,R_\mathbf{p},E,O)\right] \tag{28}$$

whenever $R_\mathbf{p}$ and P_α have the same means. The same line of reasoning applies to $P_{\alpha'}$ with the same means as P_α. Hence,

$$\begin{aligned}\mathbb{E}_{E\sim Q,O\sim P_\alpha}\left[s(Q,P_\alpha,E,O)\right] &= \mathbb{E}_{E\sim Q,O\sim R_\mathbf{p}}\left[s(Q,R_\mathbf{p},E,O)\right]\\ &= \mathbb{E}_{E\sim Q,O\sim P_{\alpha'}}\left[s(Q,P_{\alpha'},E,O)\right],\end{aligned}$$

as claimed. □

B.3 Proof of Lemma 3.3

Lemma 3.3. *Let s be a proper DSR, $Q\in\Delta(H)$ and $P_\alpha,\hat{P}_\alpha\in\Delta(\Omega)^H$ be s.t. for all $e\in H$*

$$\min_{\omega\in\Omega} u(\omega) < \mathbb{E}_{O\sim P_\alpha}\left[u(O)\mid e\right] = \mathbb{E}_{O\sim \hat{P}_\alpha}\left[u(O)\mid e\right] < \max_{\omega\in\Omega} u(\omega) \tag{4}$$

and $\operatorname{supp}(P_\alpha(\cdot\mid e))\subseteq\operatorname{supp}(\hat{P}_\alpha(\cdot\mid e))$. *Then*

$$\mathbb{E}_{E\sim Q,O\sim P_\alpha}\left[s(Q,P_\alpha,E,O)\right]=\mathbb{E}_{E\sim Q,O\sim P_\alpha}\left[s(Q,\hat{P}_\alpha,E,O)\right]. \tag{5}$$

Proof. For $p\in(0,1]$ consider

$$P'_\alpha = \frac{1}{1-p}\left(\hat{P}_\alpha - pP_\alpha\right). \tag{29}$$

Because $\operatorname{supp}(P_\alpha(\cdot\mid e))\subseteq\operatorname{supp}(\hat{P}_\alpha(\cdot\mid e))$ for all $e\in H$, there is $p\in(0,1]$ so small that $\hat{P}_\alpha(\omega\mid e) - pP_\alpha(\omega\mid e)$ is positive for all $\omega\in\Omega, e\in H$. Choose such a p for the rest of this proof. Dividing by $1-p$ renormalizes such that $P'_\alpha\in\Delta(\Omega)^H$ with

$$\hat{P}_\alpha = pP_\alpha + (1-p)P'_\alpha. \tag{30}$$

Note that for all e, $\mathbb{E}_{O\sim P'_\alpha}\left[u(O)\mid e\right]=\mathbb{E}_{O\sim\hat{P}_\alpha}\left[u(O)\mid e\right]=\mathbb{E}_{O\sim P_\alpha}\left[u(O)\mid e\right]$. That is, $P'_\alpha, P_\alpha, \hat{P}_\alpha$ all predict the same expected utility for each e.

Then

$$
\begin{aligned}
& \mathbb{E}_{\hat{P}_\alpha}\left[s(Q,\hat{P}_\alpha,E,O)\right] && (31)\\
= \quad & p\mathbb{E}_{P_\alpha}\left[s(Q,\hat{P}_\alpha,E,O)\right]+(1-p)\mathbb{E}_{P'_\alpha}\left[s(Q,\hat{P}_\alpha,E,O)\right] && (32)\\
\underset{s \text{ is proper}}{\leq} \quad & p\mathbb{E}_{P_\alpha}\left[s(Q,\hat{P}_\alpha,E,O)\right]+(1-p)\mathbb{E}_{P'_\alpha}\left[s(Q,P'_\alpha,E,O)\right] && (33)\\
\underset{s \text{ is proper}}{\leq} \quad & p\mathbb{E}_{P_\alpha}\left[s(Q,P_\alpha,E,O)\right]+(1-p)\mathbb{E}_{P'_\alpha}\left[s(Q,P'_\alpha,E,O)\right] && (34)\\
\underset{\text{Lemma 3.2}}{=} \quad & \mathbb{E}_{\hat{P}_\alpha}\left[s(Q,\hat{P}_\alpha,E,O)\right]. && (35)
\end{aligned}
$$

Because the expression at the beginning is the same as the expression in the end, the weak inequalities in the middle must be equalities. Therefore, because $p > 0$, it must be the case that $\mathbb{E}_{P_\alpha}\left[s(Q,P_\alpha,E,O)\right] = \mathbb{E}_{P_\alpha}\left[s(Q,\hat{P}_\alpha,E,O)\right]$. □

B.4 Proof of Lemma 3.4

Lemma 3.4. *Let s be a proper DSR and $\omega_1,\omega_2 \in \Omega$ be two outcomes with $u(\omega_1) = u(\omega_2)$. Let $Q \in \Delta(H), P_\alpha \in \Delta(\Omega)^H$ be s.t. for all $e \in H$*

$$\min_{\omega\in\Omega} u(\omega) < \mathbb{E}_{O\sim P_\alpha}\left[u(O) \mid e\right] < \max_{\omega\in\Omega} u(\omega). \tag{6}$$

Further, let $\omega_1,\omega_2 \in \operatorname{supp}(\hat{P}_\alpha(\cdot \mid e))$ for some $e \in H$. Then

$$s(Q,P_\alpha,e,\omega_1) = s(Q,P_\alpha,e,\omega_2). \tag{7}$$

Proof. If $u(\omega_{1/2}) = \mathbb{E}_{O\sim P_\alpha}\left[u(O) \mid e\right]$, the result follows (almost) immediately from Lemma 3.3. Else, there exists some $\omega_3 \in \operatorname{supp}(P_\alpha(\cdot \mid e))$ and $p \in (0,1]$ s.t.

$$
\begin{aligned}
O_1 \mid e &= p * \omega_1 + (1-p) * \omega_3 && (36)\\
O_2 \mid e &= p * \omega_2 + (1-p) * \omega_3 && (37)
\end{aligned}
$$

both have the same mean as $P_\alpha(\cdot \mid e)$. Further, for $e' \neq e$ let $O_{1/2} \mid e'$ be distributed according to $P_\alpha(\cdot \mid e')$. Let P_{O_i} be the conditional distribution of O_i. Then

$$
\begin{aligned}
& Q(e)ps(Q,P_\alpha,e,\omega_1) + Q(e)(1-p)s(Q,P_\alpha,e,\omega_3) + (1-Q(e))\mathbb{E}\left[s(Q,P_\alpha,E,O) \mid E \neq e\right]\\
= \quad & Q(e)\mathbb{E}\left[s(Q,P_\alpha,e,O_1) \mid e\right] + (1-Q(e))\mathbb{E}\left[s(Q,P_\alpha,E,O_1) \mid E \neq e'\right]\\
= \quad & \mathbb{E}\left[s(Q,P_\alpha,e,O_1)\right]\\
\underset{\text{Lemma 3.3}}{=} \quad & \mathbb{E}\left[s(Q,P_{O_1},e,O_1)\right]\\
\underset{\text{Lemma 3.2}}{=} \quad & \mathbb{E}\left[s(Q,P_{O_2},e,O_2)\right]\\
\underset{\text{Lemma 3.3}}{=} \quad & \mathbb{E}\left[s(Q,P_\alpha,e,O_2)\right]\\
= \quad & Q(e)\mathbb{E}\left[s(Q,P_\alpha,e,O_2) \mid e\right] + (1-Q(e))\mathbb{E}\left[s(Q,P_\alpha,E,O_2) \mid E \neq e\right]\\
= \quad & Q(e)ps(Q,P_\alpha,e,\omega_2) + Q(e)(1-p)s(Q,P_\alpha,e,\omega_3) + (1-Q(e))\mathbb{E}\left[s(Q,P_\alpha,E,O) \mid E \neq e\right]
\end{aligned}
$$

Since, $Q(e), p > 0$, it follows that $s(Q,P_\alpha,e,\omega_1) = s(Q,P_\alpha,e,\omega_2)$ as claimed. □

C Proofs for Section 4

C.1 Proof of Lemma 4.1

Lemma 4.1. *Let s be a proper DSR. Then there are functions $f_Y : \Delta(H) \times \mathbb{R}^H \to \mathbb{R}^H_{\geq 0}$, $f_E : \Delta(H) \times \mathbb{R}^H \to \mathbb{R}^H$, $g : \Delta(H) \times \mathbb{R}^H \to \mathbb{R}$, s.t. for all $\hat{Q}, \hat{\boldsymbol{\mu}}, e, y$,*

$$s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, y) = f_Y(\hat{Q}, \hat{\boldsymbol{\mu}})_e y + f_E(\hat{Q}, \hat{\boldsymbol{\mu}})_e + g(\hat{Q}, \hat{\boldsymbol{\mu}}).$$

Proof. Fix any $\hat{Q}, \hat{\boldsymbol{\mu}}, e$. We will show that $s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, \cdot)$ is affine. Specifically we show this by showing that for any random variable X over $\mathbb{R}$ with mean x it is

$$\mathbb{E}_X \left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, X) \right] = s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, x). \tag{38}$$

From this, affinity follows immediately.

So take any variable X with mean x. Let $E \sim \hat{Q}$. Further, define new random variables $Y, \tilde{Y}$ with $Y|e = p * X + (1-p) * x'$, $\tilde{Y}|e = p * x + (1-p) * x'$, where $x' \in \mathbb{R}, p \in (0,1]$ are chosen such that $\hat{Q}(e)\mathbb{E}[Y \mid e] = \hat{Q}(e)\mathbb{E}\left[\tilde{Y} \mid e\right] = \hat{\mu}_e$. For $e' \neq e$, let $Y \mid e'$ and $\tilde{Y} \mid e'$ be equally distributed with mean $\hat{\mu}_{e'}$. Then

$$\begin{aligned}
& \hat{Q}(e)p\mathbb{E}\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, X)\right] + \hat{Q}(e)(1-p)s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, x') + (1-\hat{Q}(e))\mathbb{E}\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, E, Y) \mid E \neq e\right] & (39)\\
= \ & \hat{Q}(e)\mathbb{E}\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, Y)\right] + (1-\hat{Q}(e))\mathbb{E}\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, E, Y) \mid E \neq e\right] & (40)\\
= \ & \mathbb{E}\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, Y)\right] & (41)\\
= \ & \mathbb{E}\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, \tilde{Y})\right] & (42)\\
= \ & \hat{Q}(e)\mathbb{E}\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, \tilde{Y})\right] + (1-\hat{Q}(e))\mathbb{E}\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, E, \tilde{Y}) \mid E \neq e\right] & (43)\\
= \ & \hat{Q}(e)p\mathbb{E}\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, x)\right] + \hat{Q}(e)(1-p)s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, x') + (1-\hat{Q}(e))\mathbb{E}\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, E, \tilde{Y}) \mid E \neq e\right] & (44)
\end{aligned}$$

Now, $\mathbb{E}\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, X)\right] = \mathbb{E}\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, x)\right]$ follows directly from $\hat{Q}(e) > 0, p > 0$ and

$$\mathbb{E}\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, E, Y) \mid E \neq e\right] = \mathbb{E}\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, E, \tilde{Y}) \mid E \neq e\right].$$

We have now shown that for fixed $\hat{Q}, \hat{\boldsymbol{\mu}}, e$, the function $s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, y)$ is affine in y. This means that there are functions f_Y, f_E s.t. $s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, y) = f_Y(\hat{Q}, \hat{\boldsymbol{\mu}})_e y + f_E(\hat{Q}, \hat{\boldsymbol{\mu}})_e$, as claimed (setting $g = 0$).

Finally, notice that for propriety f_Y must be non-negative – otherwise the expert would be incentivized to recommend an action that *minimizes* expected utility. □

C.2 Proof of Corollary 4.2

Corollary 4.2. *Let s be a proper scoring rule specified via f_Y, f_E, g as per Lemma 4.1. Then for all reports $\hat{Q}, \hat{\boldsymbol{\mu}}$ evidence variables E distributed according to Q and all means Y with true means $\boldsymbol{\mu}$,*

$$\mathbb{E}\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, E, Y)\right] = (f_E(\hat{Q}, \hat{\boldsymbol{\mu}}), f_Y(\hat{Q}, \hat{\boldsymbol{\mu}}))(Q, \boldsymbol{\mu}) + g(\hat{Q}, \hat{\boldsymbol{\mu}}).$$

Proof.

$$\begin{aligned}
\mathbb{E}\left[s(\hat{Q},\hat{\boldsymbol{\mu}},E,Y)\right] &= \sum_{e\in H} Q(e)\mathbb{E}\left[s(\hat{Q},\hat{\boldsymbol{\mu}},e,Y)\mid e\right] \\
&= \sum_{e\in H} Q(e)\mathbb{E}\left[f_Y(\hat{Q},\hat{\boldsymbol{\mu}})_e Y + f_E(\hat{Q},\hat{\boldsymbol{\mu}})_e + g(\hat{Q},\hat{\boldsymbol{\mu}})\mid e\right] \\
&= \sum_{e\in H} Q(e)\mathbb{E}\left[f_Y(\hat{Q},\hat{\boldsymbol{\mu}})_e Y\mid e\right] + Q(e)f_E(\hat{Q},\hat{\boldsymbol{\mu}})_e + Q(e)g(\hat{Q},\hat{\boldsymbol{\mu}}) \\
&= g(\hat{Q},\hat{\boldsymbol{\mu}}) + \sum_{e\in H} f_Y(\hat{Q},\hat{\boldsymbol{\mu}})_e\mathbb{E}\left[Q(e)Y\mid e\right] + f_E(\hat{Q},\hat{\boldsymbol{\mu}})_e Q(e) \\
&= g(\hat{Q},\hat{\boldsymbol{\mu}}) + \sum_{e\in H} f_Y(\hat{Q},\hat{\boldsymbol{\mu}})_e\mu_e + f_E(\hat{Q},\hat{\boldsymbol{\mu}})_e Q(e) \\
&= (f_E(\hat{Q},\hat{\boldsymbol{\mu}}), f_Y(\hat{Q},\hat{\boldsymbol{\mu}}))(Q,\boldsymbol{\mu}) + g(\hat{Q},\hat{\boldsymbol{\mu}})
\end{aligned}$$

□

D Proofs for Section 5

D.1 Proof of Theorem 5.1

Theorem 5.1. *Let s be a DSR. Then s is proper if and only if there exist functions $f\colon \Delta(H_{-i})\times\mathbb{R}^{H_{-i}}\to\mathbb{R}^{H_{-i}}\times\mathbb{R}^{H_{-i}}_{\geq 0}$ and $F\colon\Delta(H_{-i})\times\mathbb{R}^{H_{-i}}\to\mathbb{R}$ s.t.*

$$s(\hat{Q},\hat{\boldsymbol{\mu}},Q,\boldsymbol{\mu}) = f(\hat{Q},\hat{\boldsymbol{\mu}})((Q,\boldsymbol{\mu})-(\hat{Q},\hat{\boldsymbol{\mu}})) + F(\hat{Q},\hat{\boldsymbol{\mu}}), \tag{9}$$

where F is convex and f is a subgradient of F.

Proof. $\Leftarrow$: We first show that scoring rules of the given form are indeed proper. Because f is non-negative in those entries that are multiplied by $\boldsymbol{\mu}$, it is immediately obvious that – whatever $(\hat{Q},\hat{\boldsymbol{\mu}})$ is reported – the expert always weakly prefers reporting an optimal set of recommendations. Next, let $(Q,\boldsymbol{\mu})$ be any true evidence distribution and means and $(\hat{Q},\hat{\boldsymbol{\mu}})$ be any report. Then

$$\begin{aligned}
s(\hat{Q},\hat{\boldsymbol{\mu}},Q,\boldsymbol{\mu}) &= f(\hat{Q},\hat{\boldsymbol{\mu}})((Q,\boldsymbol{\mu})-(\hat{Q},\hat{\boldsymbol{\mu}})) + F(\hat{Q},\hat{\boldsymbol{\mu}}) && (45)\\
&\underset{\text{Ineq. 15}}{\leq} F(Q,\boldsymbol{\mu}) && (46)\\
&= F(Q,\boldsymbol{\mu}) + f(Q,\boldsymbol{\mu})((Q,\boldsymbol{\mu})-(Q,\boldsymbol{\mu})) && (47)\\
&= s(Q,\boldsymbol{\mu},Q,\boldsymbol{\mu}). && (48)
\end{aligned}$$

$\Rightarrow$: By Corollary 4.2, there are functions f, g s.t.

$$s(\hat{Q},\hat{\boldsymbol{\mu}},Q,\boldsymbol{\mu}) = f(\hat{Q},\hat{\boldsymbol{\mu}})(Q,\boldsymbol{\mu}) + g(\hat{Q},\hat{\boldsymbol{\mu}}) = f(\hat{Q},\hat{\boldsymbol{\mu}})(\hat{Q},\hat{\boldsymbol{\mu}}) + g(\hat{Q},\hat{\boldsymbol{\mu}}) + f(\hat{Q},\hat{\boldsymbol{\mu}})((Q,\boldsymbol{\mu})-(\hat{Q},\hat{\boldsymbol{\mu}})). \tag{49}$$

Now, define $F(\hat{Q},\hat{\boldsymbol{\mu}}) = f(\hat{Q},\hat{\boldsymbol{\mu}})(\hat{Q},\hat{\boldsymbol{\mu}}) + g(\hat{Q},\hat{\boldsymbol{\mu}})$. It is left to show that for s to be proper, f must be a subgradient function of F – the convexity of F follows from Lemma A.1.

For all $(Q,\boldsymbol{\mu})$ and $(\hat{Q},\hat{\boldsymbol{\mu}})$, it is

$$F(Q,\boldsymbol{\mu}) = s(Q,\boldsymbol{\mu},Q,\boldsymbol{\mu}) \underset{s\text{ proper}}{\geq} s(\hat{Q},\hat{\boldsymbol{\mu}},Q,\boldsymbol{\mu}) = F(\hat{Q},\hat{\boldsymbol{\mu}}) + f(\hat{Q},\hat{\boldsymbol{\mu}})((Q,\boldsymbol{\mu})-(\hat{Q},\hat{\boldsymbol{\mu}})). \tag{50}$$

This is exactly the subgradient inequality (Ineq. 15). □

D.2 Direct proof of Theorem 5.2

Theorem 5.2. *Let s be a DSR. Then s is proper if and only if there is a cyclically monotone increasing function $f\colon \Delta(H)\times\mathbb{R}^H \to \mathbb{R}^H \times \mathbb{R}^H_{\geq 0}$, $C\in\mathbb{R}, \mathbf{b}\in\Delta(H)\times\mathbb{R}^H$ s.t.*

$$s(\hat{Q},\hat{\boldsymbol{\mu}},Q,\boldsymbol{\mu}) = f(\hat{Q},\hat{\boldsymbol{\mu}})((Q,\boldsymbol{\mu})-(\hat{Q},\hat{\boldsymbol{\mu}})) + \int_{\mathbf{b}}^{(\hat{Q},\hat{\boldsymbol{\mu}})} f(\mathbf{z})d\mathbf{z} + C. \tag{10}$$

This section is dedicated to a direct proof of Theorem 5.2, without using any of the previous results. While one of the components of the proof is somewhat of an arithmetic grind, we think that the proof illustrates well how f uniquely determines g.

The key is the following lemma, which shows how the rate of change of g is related to the rate of change of f.

Lemma D.1. *Let $s(\hat{Q},\hat{\boldsymbol{\mu}},Q,\boldsymbol{\mu}) = f(\hat{Q},\hat{\boldsymbol{\mu}})(Q,\boldsymbol{\mu}) - g(\hat{Q},\hat{\boldsymbol{\mu}})$ be an expectation of a DSR. Then s is proper if and only if for all $(Q_a,\boldsymbol{\mu}_a),(Q_b,\boldsymbol{\mu}_b)$,*

$$(f(Q_a,\boldsymbol{\mu}_a)-f(Q_b,\boldsymbol{\mu}_b))(Q_b,\boldsymbol{\mu}_b) \leq g(Q_a,\boldsymbol{\mu}_a)-g(Q_b,\boldsymbol{\mu}_b) \leq (f(Q_a,\boldsymbol{\mu}_a)-f(Q_b,\boldsymbol{\mu}_b))(Q_a,\boldsymbol{\mu}_a). \tag{51}$$

Further, s is strictly proper if and only if Ineq. 51 is strict (in both directions) whenever $\boldsymbol{\mu}_a \neq \boldsymbol{\mu}_b$.

Proof. DSR s is proper if and only if $s(Q_a,\boldsymbol{\mu}_a,Q_b,\boldsymbol{\mu}_b) \leq s(Q_b,\boldsymbol{\mu}_b,Q_b,\boldsymbol{\mu}_b)$. This is equivalent to $f(Q_a,\boldsymbol{\mu}_a)(Q_b,\boldsymbol{\mu}_b) - g(Q_a,\boldsymbol{\mu}_a) \leq f(Q_b,\boldsymbol{\mu}_b)(Q_b,\boldsymbol{\mu}_b) - g(Q_b,\boldsymbol{\mu}_b)$, which in turn is equivalent to

$$(f(Q_a,\boldsymbol{\mu}_a)-f(Q_b,\boldsymbol{\mu}_b))(Q_b,\boldsymbol{\mu}_b) \leq g(Q_a,\boldsymbol{\mu}_a)-g(Q_b,\boldsymbol{\mu}_b). \tag{52}$$

The other inequality is similarly equivalent to $s(Q_b,\boldsymbol{\mu}_b,Q_a,\boldsymbol{\mu}_a) \leq s(Q_a,\boldsymbol{\mu}_a,Q_a,\boldsymbol{\mu}_a)$. □

Lemma D.2. *Let $f\colon\mathbb{R}^n\to\mathbb{R}^n, g\colon\mathbb{R}^n\to\mathbb{R}$. If Ineq. 51 holds for all $(Q_a,\boldsymbol{\mu}_a),(Q_b,\boldsymbol{\mu}_b)$, then f is cyclically monotone increasing.*

Proof. Let $(Q_1,\boldsymbol{\mu}_1),...,(Q_n,\boldsymbol{\mu}_n),(Q_{n+1},\boldsymbol{\mu}_{n+1})=(Q_1,\boldsymbol{\mu}_1)$. Then

$$\sum_{i=1}^n (f(Q_{i+1},\boldsymbol{\mu}_{i+1})-f(Q_i,\boldsymbol{\mu}_i))(Q_i,\boldsymbol{\mu}_i) \underset{\text{Lemma D.1}}{\leq} \sum_{i=1}^n g(Q_{i+1},\boldsymbol{\mu}_{i+1})-g(Q_i,\boldsymbol{\mu}_i) = 0,$$

as required. □

The idea now is for given f, Ineq. 51 specifies g uniquely up to a constant, by considering $(Q_a,\boldsymbol{\mu}_a)$ and $(Q_b,\boldsymbol{\mu}_b)$ that are infinitesimally close to one another.

Lemma D.3. *Let f be cyclically monotone increasing. Then the set of functions defined by*

$$g(\hat{Q},\hat{\boldsymbol{\mu}}) = f(\hat{Q},\hat{\boldsymbol{\mu}})(\hat{Q},\hat{\boldsymbol{\mu}}) - \int_{\mathbf{b}}^{(\hat{Q},\hat{\boldsymbol{\mu}})} f(\mathbf{x})d\mathbf{x} + C, \tag{53}$$

for any $C\in\mathbb{R},\mathbf{b}\in\Delta(H)\times\mathbb{R}^H$ are exactly the functions that satisfy Ineq. 51 for all $(Q_a,\boldsymbol{\mu}_a),(Q_b,\boldsymbol{\mu}_b)$.

Proof. $\Leftarrow$: First we show that if g satisfies Ineq. 51 for given f, g must be of the form in Eq. 53.

Fix any $(\hat{Q},\hat{\boldsymbol{\mu}})$. For $n\in\mathbb{N}$, let $\mathbf{x}_{n,0},\mathbf{x}_{n,1},...,\mathbf{x}_{n,n}$ be in $[\mathbf{b},(\hat{Q},\hat{\mu})]$. Let these be ordered in the natural way with $\mathbf{x}_{n,0}=\mathbf{b}$ and $\mathbf{x}_{n,n}=(\hat{Q},\hat{\boldsymbol{\mu}})$. For example, we could let $\mathbf{x}_{n,i}=\mathbf{x}_{n,i-1}+((\hat{Q},\hat{\mu})-\mathbf{b})/n$ By telescoping, we can write:

$$g(\hat{Q},\hat{\boldsymbol{\mu}}) = g(\mathbf{b}) + \sum_{i=1}^n g(\mathbf{x}_{n,i}) - g(\mathbf{x}_{n,i-1}). \tag{54}$$

Since relative to any f, g can only be unique up to a constant, we will write C instead of $g(\mathbf{b})$. From Lemma D.1, it follows that

$$\begin{aligned} & \sum_{i=1}^{n} \mathbf{x}_{n,i-1} \left(f\left(\mathbf{x}_{n,i}\right) - f\left(\mathbf{x}_{n,i-1}\right)\right) && (55) \\ \leq \; & g(\hat{Q}, \hat{\boldsymbol{\mu}}) - C && (56) \\ \leq \; & \sum_{i=1}^{n} \mathbf{x}_{n,i} \left(f\left(\mathbf{x}_{n,i}\right) - f\left(\mathbf{x}_{n,i-1}\right)\right) && (57) \end{aligned}$$

for all $n \in \mathbb{N}_{>0}$.

We would now like to find g by taking the limit w.r.t. $n \to \infty$ of the two series, where we let the partitions $(\mathbf{x}_{n,i})_i$ become arbitrarily fine as $n \to \infty$. To do so, we will rewrite the two sums to interpret them as the (right and left) Riemann sums of some function.[3] It is

$$\begin{aligned} & \sum_{i=1}^{n} \mathbf{x}_{n,i} \left(f\left(\mathbf{x}_{n,i}\right) - f\left(\mathbf{x}_{n,i-1}\right)\right) && (58) \\ = \; & \sum_{i=1}^{n} \mathbf{x}_{n,i} f\left(\mathbf{x}_{n,i}\right) - \mathbf{x}_{n,i-1} f\left(\mathbf{x}_{n,i-1}\right) - \sum_{i=1}^{n} (\mathbf{x}_{n,i} - \mathbf{x}_{n,i-1}) f\left(\mathbf{x}_{n,i-1}\right) && (59) \\ = \; & (\hat{Q}, \hat{\boldsymbol{\mu}}) f(\hat{Q}, \hat{\boldsymbol{\mu}}) - \mathbf{b} f(\mathbf{b}) - \sum_{i=1}^{n} (\mathbf{x}_{n,i} - \mathbf{x}_{n,i-1}) f\left(\mathbf{x}_{n,i-1}\right). && (60) \end{aligned}$$

The last step is due to telescoping of the left-hand sum. Analogously,

$$\begin{aligned} & \sum_{i=1}^{n} \mathbf{x}_{n,i-1} \left(f\left(\mathbf{x}_{n,i}\right) - f\left(\mathbf{x}_{n,i-1}\right)\right) && (61) \\ = \; & \sum_{i=1}^{n} \mathbf{x}_{n,i} f\left(\mathbf{x}_{n,i}\right) - \mathbf{x}_{n,i-1} f\left(\mathbf{x}_{n,i-1}\right) - \sum_{i=1}^{n} (\mathbf{x}_{n,i} - \mathbf{x}_{n,i-1}) f\left(\mathbf{x}_{n,i}\right) && (62) \\ = \; & (\hat{Q}, \hat{\boldsymbol{\mu}}) f(\hat{Q}, \hat{\boldsymbol{\mu}}) - \mathbf{b} f(\mathbf{b}) - \sum_{i=1}^{n} (\mathbf{x}_{n,i} - \mathbf{x}_{n,i-1}) f\left(\mathbf{x}_{n,i}\right). && (63) \end{aligned}$$

By Lemma A.5,

$$\sum_{i=1}^{n} (\mathbf{x}_{n,i} - \mathbf{x}_{n,i-1}) f\left(\mathbf{x}_{n,i-1}\right) \underset{k\to\infty}{\to} \int_{0}^{(\hat{Q}, \hat{\boldsymbol{\mu}})} f(\mathbf{x}) d\mathbf{x} \underset{n\to\infty}{\leftarrow} \sum_{i=1}^{n} (\mathbf{x}_{n,i} - \mathbf{x}_{n,i-1}) f\left(\mathbf{x}_{n,i}\right). \qquad (64)$$

So for $k \to \infty$, the lower and upper bound on $g(\hat{Q}, \hat{\mu})$ converge to the same value. Hence, $g(\hat{Q}, \hat{\mu})$ must be that value, i.e.

$$g(\hat{Q}, \hat{\mu}) = C + (\hat{Q}, \hat{\mu}) f(\hat{Q}, \hat{\mu}) - \int_{0}^{\hat{Q}, \hat{\mu}} f(x) dx, \qquad (65)$$

as claimed.

[3] In fact, we could immediately interpret them as Riemann sums of the function f^{-1}, see Appendix A.5 and in particular the proof of Theorem A.8.

$\Rightarrow$: Consider any $(\hat{Q}_a, \hat{\boldsymbol{\mu}}_a)$ and $(\hat{Q}_b, \hat{\boldsymbol{\mu}}_b)$. It is

$$\begin{aligned}
& g(Q_a, \boldsymbol{\mu}_a) - g(Q_b, \boldsymbol{\mu}_b) \\
= \; & f(Q_a, \boldsymbol{\mu}_a)(Q_a, \boldsymbol{\mu}_a) - \int_{\mathbf{b}}^{(Q_a, \boldsymbol{\mu}_a)} f(\mathbf{x}) d\mathbf{x} - f(Q_b, \boldsymbol{\mu}_b)(Q_b, \boldsymbol{\mu}_b) + \int_{\mathbf{b}}^{(Q_b, \boldsymbol{\mu}_b)} f(\mathbf{x}) d\mathbf{x} \\
= \; & (f(Q_a, \boldsymbol{\mu}_a) - f(Q_b, \boldsymbol{\mu}_b))(Q_b, \boldsymbol{\mu}_b) - f(Q_a, \boldsymbol{\mu}_a)((Q_b, \boldsymbol{\mu}_b) - (Q_a, \boldsymbol{\mu}_a)) - \int_{(Q_b, \boldsymbol{\mu}_b)}^{(Q_a, \boldsymbol{\mu}_a)} f(\mathbf{x}) d\mathbf{x}.
\end{aligned}$$

Now let $(\mathbf{x}_{n,i})_{n,i}$ be partitions of $[(Q_b, \boldsymbol{\mu}_b), (Q_a, \boldsymbol{\mu}_a)]$ that become arbitrarily fine. Then

$$\begin{aligned}
& (f(Q_a, \boldsymbol{\mu}_a) - f(Q_b, \boldsymbol{\mu}_b))(Q_b, \boldsymbol{\mu}_b) - f(Q_a, \boldsymbol{\mu}_a)((Q_b, \boldsymbol{\mu}_b) - (Q_a, \boldsymbol{\mu}_a)) - \int_{(Q_b, \boldsymbol{\mu}_b)}^{(Q_a, \boldsymbol{\mu}_a)} f(\mathbf{x}) d\mathbf{x} \\
= \; & (f(\mathbf{x}_{n,n}) - f(\mathbf{x}_{n,0}))(\mathbf{x}_{n,0}) - f(\mathbf{x}_{n,n})(\mathbf{x}_{n,0} - \mathbf{x}_{n,n}) - \int_{\mathbf{x}_{n,0}}^{\mathbf{x}_{n,n}} f(\mathbf{x}) d\mathbf{x} \\
\underset{n \to \infty}{\leftarrow} \; & (f(\mathbf{x}_{n,n}) - f(\mathbf{x}_{n,0}))(\mathbf{x}_{n,0}) - f(\mathbf{x}_{n,n})(\mathbf{x}_{n,0} - \mathbf{x}_{n,n}) - \sum_{i=1}^{n} f(\mathbf{x}_{n,i-1})(\mathbf{x}_{n,i} - \mathbf{x}_{n,i-1}) \\
\underset{\text{cyc. mon.}}{\geq} \; & (f(\mathbf{x}_{n,n}) - f(\mathbf{x}_{n,0}))(\mathbf{x}_{n,0}) \\
= \; & (f(Q_a, \boldsymbol{\mu}_a) - f(Q_b, \boldsymbol{\mu}_b))(Q_a, \boldsymbol{\mu}_b)
\end{aligned}$$

□

Together, Lemmas D.1 to D.3 imply Theorem 5.2.

D.3 Proof of Lemma 5.3

Lemma 5.3. *Let s be a proper DSR with $s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) = f(\hat{Q}, \hat{\boldsymbol{\mu}})\boldsymbol{\mu} - g(\hat{Q}, \hat{\boldsymbol{\mu}})$ for all $\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}$. Then for all $(\hat{Q}_1, \hat{\boldsymbol{\mu}}_1), (\hat{Q}_2, \hat{\boldsymbol{\mu}}_2)$, $f(\hat{Q}_1, \hat{\boldsymbol{\mu}}_1) = f(\hat{Q}_1, \hat{\boldsymbol{\mu}}_2) \implies g(\hat{Q}_1, \hat{\boldsymbol{\mu}}_1) = g(\hat{Q}_1, \hat{\boldsymbol{\mu}}_2)$.*

Proof. If $f(\hat{Q}_1, \hat{\boldsymbol{\mu}}_1) = f(\hat{Q}_2, \hat{\boldsymbol{\mu}}_2)$, but (WLOG) $g(\hat{Q}_1, \hat{\boldsymbol{\mu}}_1) > g(\hat{Q}_2, \hat{\boldsymbol{\mu}}_2)$, then the expert would always strictly prefer reporting $(\hat{Q}_2, \hat{\boldsymbol{\mu}}_2)$ over reporting $(\hat{Q}_1, \hat{\boldsymbol{\mu}}_1)$ – even when the true evidence distribution and means are $(\hat{Q}_1, \hat{\boldsymbol{\mu}}_1)$. This contradicts the propriety of s. □

D.4 Proof of Lemma 5.4

Lemma 5.4. *Let $\tilde{g}$ be the quantity-price function of a proper DSR. Then g is convex.*

Proof. Let $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q} \in \operatorname{im}(f)$ be pairwise non-equal and $p \in (0, 1)$ s.t. $\mathbf{q} = p\mathbf{q}_1 + (1-p)\mathbf{q}_2$. Imagine for contradiction that

$$p\tilde{g}(\mathbf{q}_1) + (1-p)\tilde{g}(\mathbf{q}_2) < g(\mathbf{q}). \tag{66}$$

Then the expert's average/expected score of randomizing between buying $\mathbf{q}_1$ and $\mathbf{q}_2$ is always better than buying $\mathbf{q}$ shares regardless of what the true means $\boldsymbol{\mu}$ are:

$$\begin{aligned}
& p(\mathbf{q}_1(Q, \boldsymbol{\mu}) - \tilde{g}(\mathbf{q}_1)) + (1-p)(\mathbf{q}_2(Q, \boldsymbol{\mu}) - \tilde{g}(\mathbf{q}_2)) && (67) \\
= \; & (p\mathbf{q}_1 + (1-p)\mathbf{q}_2)(Q, \boldsymbol{\mu}) - (p\tilde{g}(\mathbf{q}_1) + (1-p)\tilde{g}(\mathbf{q}_2)) && (68) \\
= \; & \mathbf{q}(Q, \boldsymbol{\mu}) - (p\tilde{g}(\mathbf{q}_1) + (1-p)\tilde{g}(\mathbf{q}_2)) && (69) \\
\underset{\text{Ineq. 66}}{>} \; & \mathbf{q}(Q, \boldsymbol{\mu}) - \tilde{g}(\mathbf{q}). && (70)
\end{aligned}$$

If the average of two numbers is above some number, then at least one of the two former numbers is above the latter number, so for each $(Q, \boldsymbol{\mu})$, it must be

$$\mathbf{q}_1(Q, \boldsymbol{\mu}) - \tilde{g}(\mathbf{q}_1) > \mathbf{q}(Q, \boldsymbol{\mu}) - \tilde{g}(\mathbf{q}) \text{ or } \mathbf{q}_2(Q, \boldsymbol{\mu}) - \tilde{g}(\mathbf{q}_2) > \mathbf{q}(Q, \boldsymbol{\mu}) - \tilde{g}(\mathbf{q}). \tag{71}$$

This means that the expert never chooses to buy the quantity $\mathbf{q}$, in contradiction to $\mathbf{q} \in \operatorname{im}(f)$. □

E A generalization of Othman and Sandholm's (2010) characterization

Consider the special case of differentiable DSRs. Since the subgradient is equal to the gradient for differentiable convex functions, we could directly obtain characterizations of differentiable scoring rules from Theorems 5.1 and 5.5. However, we here give a third characterization of differentiable proper DSRs, which generalizes the characterization of differentiable proper DSRs for the case $|H| = 1, |\Omega| = 2$ of Othman and Sandholm (2010, Section 2.3.2).

Proposition E.1. *Let s be a differentiable DSR. Then s is proper if and only if there are cyc. mon. incr., differentiable $f\colon \Delta(H) \times \mathbb{R}^H \to \mathbb{R}^H \times \mathbb{R}^H_{\geq 0}$ and differentiable $g\colon \mathbb{R}^H \to \mathbb{R}$ s.t.*

$$s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) = f(\hat{Q}, \hat{\boldsymbol{\mu}})(Q, \boldsymbol{\mu}) - g(\hat{Q}, \hat{\boldsymbol{\mu}}) \tag{72}$$

with

$$(Q, \boldsymbol{\mu}) \frac{d}{d\mu_e} f(Q, \boldsymbol{\mu}) = \frac{d}{d\mu_e} g(Q, \boldsymbol{\mu}) \tag{73}$$

$$(Q, \boldsymbol{\mu}) \frac{d}{dQ(e)} f(Q, \boldsymbol{\mu}) = \frac{d}{dQ(e)} g(Q, \boldsymbol{\mu}) \tag{74}$$

for all $e \in H, (Q, \boldsymbol{\mu}) \in \Delta(\Omega) \times \mathbb{R}^H$.

Proof. $\Rightarrow$: By Corollary 4.2, there are functions f, g s.t. the expected scores are $s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) = f(\hat{Q}, \hat{\boldsymbol{\mu}})\boldsymbol{\mu} - g(\hat{Q}, \hat{\boldsymbol{\mu}})$. First, we have already shown that f must be cyclically monotone (Lemma D.2). Note that if $s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, y)$ is differentiable, so is the expected score $s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu})$, because the linear combination of differentiable functions is differentiable. Also, if $s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu})$ is everywhere differentiable, f, g must be everywhere differentiable.

For s to be proper, for each $(Q, \boldsymbol{\mu})$, $s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu})$ has to be maximal at $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}$. Hence, for each e, $\frac{d}{d\hat{\mu}_e} s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) = 0$ at $\boldsymbol{\mu} = \hat{\boldsymbol{\mu}}$. Now,

$$\begin{aligned} \frac{d}{d\hat{\mu}_e} s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) &= (Q, \boldsymbol{\mu}) \frac{d}{d\hat{\mu}_e} f(\hat{Q}, \hat{\boldsymbol{\mu}}) - \frac{d}{d\hat{\mu}_e} g(\hat{Q}, \hat{\boldsymbol{\mu}}) \\ \frac{d}{d\hat{Q}(e)} s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) &= (Q, \boldsymbol{\mu}) \frac{d}{d\hat{Q}(e)} f(\hat{Q}, \hat{\boldsymbol{\mu}}) - \frac{d}{d\hat{Q}(e)} g(\hat{Q}, \hat{\boldsymbol{\mu}}). \end{aligned}$$

For this to be zero at $(\hat{Q}, \hat{\boldsymbol{\mu}}) = (Q, \boldsymbol{\mu})$, Equations 73 and 74 must hold, as claimed.

$\Leftarrow$: For the other direction, we provide a short but indirect argument. Notice first that any cyclically monotonically increasing, differentiable f the derivatives of f, which in turn determine uniquely via Equations 73 and 74 the derivatives of g, which in turn determine g uniquely up to a constant. In sum, any cyclically monotonically increasing, differentiable f of the given type signature determines a set of scoring rules that satisfy Equations 73 and 74 and that differ only by a constant. Call this set of scoring rules M_f. By the proof of the other direction of the proposition above ($\Rightarrow$), all proper DSRs with the given f are in this set. By our other characterization there are proper DSRs for the given (cyclically monotonically increasing) f (whose f_Y entries are positive). Thus, M_f contains at least one proper DSR. Since any pair of scoring rules in M_f differ only by a constant and adding a constant preserves propriety, all scoring rules in M_f are proper. □

We briefly show this is indeed equivalent to the characterization of Othman and Sandholm (2010) in the $|H| = 1, |\Omega| = 2$ special case. They characterize such differentiable proper scoring rules s as ones where A) $s(p, 1) > s(p, 0)$ and B) $\frac{s'(p,1)}{s'(p,0)} = \frac{p-1}{p}$ for all p. A is equivalent to being able to write

$s(p,1) = f(p) - g(p)$ and $s(p,0) = -g(p)$ for some differentiable g and positive, differentiable f. With that we can re-write B:

$$\frac{s'(p,1)}{s'(p,0)} = \frac{p-1}{p} \Leftrightarrow \frac{f'(p) - g'(p)}{g'(p)} = \frac{p-1}{p} \Leftrightarrow g'(p) = pf'(p), \tag{75}$$

which is exactly the relationship between f', g' stated in our Proposition E.1. Othman and Sandholm seem to forget the necessity of $s'(p,1) > 0$ (i.e., the monotonicity of s). Note that Othman and Sandholm use different names for scoring rules. In particular, they use "$f(p)$" for $s(p,1)$ and "$g(p)$" for $s(p,0)$.

Can Proposition E.1 be generalized to *non*-differentiable DSRs s, giving yet another style of characterization? At least on first sight, this seems difficult, because if f could be discontinuous and non-convex, making it unclear what concept could replace the derivative of f.