# Decision Scoring Rules

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

**Abstract**

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: 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 the design of mechanisms for eliciting recommendations and predictions from multiple experts. With a few modifications, the above characterization for the single-expert case carries over. Among other things, this characterization implies that the mechanism should not allow experts 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 probabilistic 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? These are the main questions of this paper.

Our approach to formalizing and answering our main 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 $\omega$, 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; cf. Boutilier, 2012). 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, we let the expert's score depend only on the reported expected utility for the recommended action. Next we show that the scoring rule must be affine in utility of the outcome obtained. Using these results, we then obtain five different (equivalent) characterizations of proper decision scoring rules (Section 5). The first two (Section 5.1) are analogous to existing results on proper affine scoring (Frongillo and Kash, 2014), including the characterizations of proper scoring rules by Gneiting and Raftery (2007). The other three characterizations (Section 5.2), which we have not seen anywhere in this form for affine scoring, have 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 6, we discuss the implications of our characterization for mechanism for eliciting decision-relevant mechanisms from multiple experts. Finally, we discuss related work in Section 7.

[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.

## 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 $\Omega$. The principal would like to choose an action that maximizes the expectation of a utility function $u\colon \Omega \to \mathbb{R}$. We assume $|u(\Omega)| \geq 2$, i.e., that the principal indeed has a strict preference over outcomes. 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 knows neither the distributions over $\Omega$ 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 probability distribution $Q \in \Delta(H)$ over $H$. Furthermore, for each $e \in H$ and $a \in A$, he believes that the outcome will be distributed according to some $P(\cdot \mid e, a) \in \Delta(\Omega)$ if $e$ is observed and action $a$ is taken. The principal asks the expert to submit any *report* $(\hat{\alpha}, \hat{Q}, \hat{P}_\alpha)$, where $\alpha \in A^H$ specifies for each $e \in H$ specifies a *recommendation* $\alpha(e) \in A$, $\hat{Q} \in \Delta(H)$ is a probabilistic prediction of what private evidence the principal observes, and $\hat{P}_\alpha \in \Delta(\Omega)^H$ provides for each $e \in H$ a probabilistic prediction over what outcome will materialize once the principal has observed any $e$ and chosen the recommended $\alpha(e)$.

The principal would like the expert to reveal these truthfully. Formally, we call a recommendation function $\alpha$ truthful if it is optimal as judged by the principal's utility function and the expert's probabilistic beliefs, i.e., if for all $e \in H$, $\alpha(e) \in \arg\max_{a\in A} \sum_{\omega\in\Omega} P(\omega \mid e, a)u(\omega)$. The conditional outcome prediction $\hat{P}_\alpha$ is truthful if for all $e \in H$, $P_\alpha(\omega \mid e) = P(\omega \mid e, \alpha(e))$. Finally, we call the evidence prediction $\hat{Q}$ truthful if $\hat{Q} = Q$. Note that, because there might be multiple optimal actions for some $e \in H$, there are multiple truthful reports. As usual, the principal cannot verify directly whether a report is truthful.

Once the expert has submitted his report and the principal has observed $E = e \in H$, the principal chooses the recommended action $\alpha(e)$. He then observes an outcome $\omega$.

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 $\omega$ 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

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

for some honest recommendation $\alpha^*$.

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 (Theorem 3.1), 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 $\alpha$, there exists an honest report $(\alpha^*, Q, P_{\alpha^*})$ s.t.

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

**Example 1** (Linear decision 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 for each $e$, the principal gives the expert $c_e (\geq 0)$ assets that each pay, say, \$1 per unit of utility obtained if $e$ is observed by the principal. If the principal is a firm, then these assets can be seen as shares in the company. However, each of these shares only pays conditional on some $e$. Therefore, in allusion to the concept of Arrow-Debreu securities, we call these assets *Arrow-Debreu shares* in the principal's project.

**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.

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

**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 $P_\alpha, \omega$ 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 prediction. 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 fully overcome.

**Example 3** (Misapplying Brier's (1950) scoring rule for outcome prediction)**.** Define the following improper DSR: $s(\hat{Q}, \hat{P}_\alpha, e, \omega) = 2\hat{P}_\alpha(\omega \mid e) - \sum_{\omega' \in \Omega} \hat{P}_\alpha(\omega' \mid e)^2$. Once again this is Brier's scoring rule. However, this time it is applied to the submitted outcome prediction, conditional on the observed $e$. Again it can be shown that to maximize his expected 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 for each $e \in H$ 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 $|H| = 1$, that $\Omega = \{\omega_1, ..., \omega_m\}$, that the optimal action $a^*$ leads to the outcome being uniformly distributed, $O_{a^*} = \frac{1}{m} * \omega_1 + ... + \frac{1}{m} * \omega_m$, 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 $\mathbb{E}\left[s(\hat{Q}, P_{a'}, E, O_{a'})\right] = 1 > \frac{1}{m} = \frac{2}{m} - \sum_{w' \in \Omega} \frac{1}{m^2} = \mathbb{E}\left[s(\hat{Q}, P_{a^*}, E, O_{a^*})\right]$.

## 3 Only expected utilities can be strictly elicited

Next, we prove that if a DSR is to be proper, it can only strictly incentivize the expert to be honest about the evidence prediction, optimal (recommended) actions and the *expected utility* of those recommended actions. As a consequence, proper DSRs in general cannot strictly incentivize the expert to honestly reveal, say, the variance of the utility under the recommended actions, or the probability of a particular outcome.

**Theorem 3.1.** *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 $e \in H$, $\min_{\omega \in \Omega} u(\omega) < \mathbb{E}_{O \sim P_\alpha}[u(O) \mid e] = \mathbb{E}_{O \sim \hat{P}_\alpha}[u(O) \mid e] < \max_{\omega \in \Omega} u(\omega)$ and $\operatorname{supp}(P_\alpha(\cdot \mid e)) \subseteq \operatorname{supp}(\hat{P}_\alpha(\cdot \mid e))$. Then $\mathbb{E}_{E \sim Q, O \sim P_\alpha}[s(Q, P_\alpha, E, O)] = \mathbb{E}_{E \sim Q, O \sim P_\alpha}\left[s(Q, \hat{P}_\alpha, E, O)\right]$.*

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 3.2.** *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 $e \in H$, $\min_{\omega\in\Omega} u(\omega) < \mathbb{E}_{O\sim P_\alpha}[u(O) \mid e] < \max_{\omega\in\Omega} u(\omega)$. Further, let $\omega_1, \omega_2 \in \operatorname{supp}(\hat{P}_\alpha(\cdot \mid e))$ for some $e \in H$. Then $s(Q, P_\alpha, e, \omega_1) = s(Q, P_\alpha, e, \omega_2)$.*

Because of Theorem 3.1 and lemma 3.2, 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, for any true belief $Q, P_\alpha$ and any $e \in H$, we call $\mu_e = \mathbb{E}_{E\sim Q,O\sim P_\alpha}[\mathbb{1}[E{=}e]\, u(O)] = Q(e)\mathbb{E}_{O\sim P_\alpha}[u(O) \mid e]$, the *Arrow-Debreu (A-D) mean* for $e$ in allusion to the concept of Arrow-Debreu shares and securities. We then want the expert to report the A-D means, rather than the regular conditional 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 A-D 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 $\hat{Q}(e) = 0$.

**Definition 4.** We say that $s$ is strictly proper w.r.t. the means if it is proper and for all beliefs $Q, P$ with A-D means $\boldsymbol{\mu}$ and all reports $(\alpha, \hat{Q}, \hat{\boldsymbol{\mu}})$ for which there is $e$ with $\hat{\mu}_e \neq \mu_e$ and $Q(e) > 0$, it is $\mathbb{E}_{E\sim Q,O\sim P_\alpha}\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, E, O)\right] < \mathbb{E}_{E\sim Q,O\sim P_\alpha}[s(Q, \boldsymbol{\mu}, E, O)]$.

# 4 Proper decision scoring rules are (non-negative) affine

We now show a result about the shape of proper DSRs. The result essentially establishes that the expected score is affine in the true A-D means $\boldsymbol{\mu}$ and the true distribution $Q$. Note that in the case of $Q$ this is trivial by the definition of the expectations, so the substantial part of the result is affinity in $\boldsymbol{\mu}$.

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

Note that $f_Y$ is non-negative, making the expected scores *non-negative* affine in $\boldsymbol{\mu}$. Clearly, this is needed to make sure that the expert makes recommendations that maximize (rather than minimize) the principal's expected utility. Note also 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 $s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) \coloneqq f(\hat{Q}, \hat{\boldsymbol{\mu}})(Q, \boldsymbol{\mu}) + g(\hat{Q}, \hat{\boldsymbol{\mu}})$, 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 that we will ignore) 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 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 5.5) to be particularly intuitive.

## 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 $f\colon \Delta(H_{-i}) \times \mathbb{R}^{H_{-i}} \to \mathbb{R}^{H_{-i}} \times \mathbb{R}^{H_{-i}}_{\geq 0}$ and $F\colon \Delta(H_{-i}) \times \mathbb{R}^{H_{-i}} \to \mathbb{R}$ s.t. $s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) = f(\hat{Q}, \hat{\boldsymbol{\mu}})((Q, \boldsymbol{\mu}) - (\hat{Q}, \hat{\boldsymbol{\mu}})) + F(\hat{Q}, \hat{\boldsymbol{\mu}})$, where $F$ is convex and $f$ is a subgradient of $F$. Further, $s$ is strictly proper w.r.t. the recommendation for $e$ if $f_{Y,e} > 0$.*

This follows directly from our affinity results in Section 4 combined with the general characterization of linear scoring rules by Frongillo and Kash (2014). As noted by Gneiting and Raftery (2007) and Frongillo and Kash (2014), we can use Theorem A.6 (and Theorem A.13) to obtain the following alternative characterization. The characterization uses the concept of cyclic monotonicity, the relevance of which for this type of characterization was first recognized by Rochet (1987).

**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.*

$$s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) = f(\hat{Q}, \hat{\boldsymbol{\mu}})((Q, \boldsymbol{\mu}) - (\hat{Q}, \hat{\boldsymbol{\mu}})) + \int_{\mathbf{b}}^{(\hat{Q}, \hat{\boldsymbol{\mu}})} f(\mathbf{z}) d\mathbf{z} + C. \tag{1}$$

The integral here is a path integral, as discussed in Appendix A. Because it is path independent (Lemma A.10), it suffices to specify the endpoints of the path ($\mathbf{b}$ and $(\hat{Q}, \hat{\boldsymbol{\mu}})$).

**Example 4.** If we assume true and reported means to always be positive, 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

By Lemma 4.1, every proper DSR can be written as $s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, y) = f_Y(\hat{Q}, \hat{\boldsymbol{\mu}})_e y + f_E(\hat{Q}, \hat{\boldsymbol{\mu}})_e + g(\hat{Q}, \hat{\boldsymbol{\mu}})$. In line with our interpretation of Example 1, we can interpret any proper DSR as follows. By submitting $(\hat{Q}, \hat{\boldsymbol{\mu}})$, the expert buys $f_Y(\hat{Q}, \hat{\boldsymbol{\mu}})_e$ Arrow-Debreu shares and $f_E(\hat{Q}, \hat{\boldsymbol{\mu}})_e$ Arrow-Debreu securities in every evidence value $e$, at an overall price of $g(\hat{Q}, \hat{\boldsymbol{\mu}})$. Unfortunately, in the characterizations of Section 5.1, the price term

$(g(\hat{Q},\hat{\boldsymbol{\mu}})f(\hat{Q},\hat{\boldsymbol{\mu}})(\hat{Q},\hat{\boldsymbol{\mu}}) - \int_{\mathbf{b}}^{(\hat{Q},\hat{\boldsymbol{\mu}})} f(\mathbf{z})d\mathbf{z} - C)$ is unintuitive. In this section, we provide a second cluster of three characterizations that correspond very intuitively to the mechanism of offering the expert A-D shares and securities at different prices. In fact, even the derivation of the characterizations (though not in its path integral form) can be comfortably put in these terms, which is why we give it in the main text.

We start with a lemma which says that the quantity of shares and securites bought $(f(\hat{Q},\hat{\boldsymbol{\mu}}))$ uniquely determines the price.

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

Instead of having the expert report $(\hat{Q},\hat{\boldsymbol{\mu}})$, we can therefore imagine that the expert chooses a quantity vector $\mathbf{q}(= f(\hat{Q},\hat{\boldsymbol{\mu}})) \in \mathrm{im}(f)$ of Arrow-Debreu shares and securities. (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 beliefs $(Q,\boldsymbol{\mu})$ 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. (Intuitively, in the offering-quantities framework if there were multiple prices for the same quantity, the expert would always choose the lowest price.) We can therefore define $\tilde{G}\colon \mathrm{im}(f) \to \mathbb{R}\colon \mathbf{q} \mapsto g(f^{-1}(\mathbf{q}))$ to be the *quantity-price function* of $s$. Instead of specifying a scoring rule $s$ in terms of a pair $f,g$, we can specify it in terms of a pair $f,\tilde{G}$ or equivalently $f^{-1},\tilde{G}$.

To characterize proper DSRs, we need to specify what property $f/f^{-1}$ and $\tilde{G}$ must (jointly) satisfy. By definition, propriety means that under any belief $(Q,\boldsymbol{\mu})$, $(Q,\boldsymbol{\mu})$ is among the optimal reports. With our interpretation of DSRs, this is equivalent to saying that the quantity $f(Q,\boldsymbol{\mu})$ is among the optimal quantities to buy under the belief $(Q,\boldsymbol{\mu})$ and under the prices specified by $\tilde{G}$. By the definition of inverses, another equivalent way of stating this relationship between $f$ and $\tilde{G}$ is the following. For any quantity vector $\mathbf{q}_0 \in \mathrm{im}(f)$, $f^{-1}(\mathbf{q}_0)$ must contain exactly those beliefs $(Q,\boldsymbol{\mu})$ under which the expert would choose to buy quantities $\mathbf{q}_0$ if the quantity-price function is $\tilde{G}$. In other words, for all $\mathbf{q}_0$, $f^{-1}(\mathbf{q}_0)$ is the set of all $(Q,\boldsymbol{\mu})$ s.t. for all $\mathbf{q} \in \mathrm{im}(f)$, $\mathbf{q}_0(Q,\boldsymbol{\mu}) - \tilde{G}(\mathbf{q}_0) \geq \mathbf{q}(Q,\boldsymbol{\mu}) - \tilde{G}(\mathbf{q})$. Now rearranging this inequality gives $\tilde{G}(\mathbf{q}) \geq \tilde{G}(\mathbf{q}_0) + f^{-1}(\mathbf{q}_0)(\mathbf{q}-\mathbf{q}_0)$, the subgradient inequality. That is, we have shown that propriety is equivalent to $f^{-1}$ being a subgradient of $\tilde{G}$. With this, we get the following characterization.

**Theorem 5.4.** *Let $s$ be a DSR. Then $s$ is proper if and only if there are* $f\colon \Delta(H)\times\mathbb{R}^H \to \mathbb{R}^H \times \mathbb{R}^H_{\geq 0}$, $\tilde{G}\colon \mathrm{im}(f) \to \mathbb{R}$ *s.t.* $s(\hat{Q},\hat{\boldsymbol{\mu}},Q,\boldsymbol{\mu}) = f(\hat{Q},\hat{\boldsymbol{\mu}})(Q,\boldsymbol{\mu}) - \tilde{G}(f(Q,\hat{\boldsymbol{\mu}}))$ *and* $f^{-1}$ *is a subgradient of* $\tilde{G}$.

Like Theorem 5.1, this theorem is stated in terms of subgradients. However, as shown in the above derivation, the subgradient inequality now directly expresses the intuitive requirement that $f^{-1}(\mathbf{q})$ infers the expert's beliefs $(Q,\boldsymbol{\mu})$ from a choice of quantity $\mathbf{q}$ under the quantity-price function $\tilde{G}$. As in Section 5.1, we can replace the function $\tilde{G}$ with subgradient $f^{-1}$ by a path integral over $f^{-1}$. We thus obtain the following characterization.

**Theorem 5.5.** *Let $s$ be a DSR. Then $s$ is (strictly) proper if and only if there is cyclically monotone increasing* $f\colon \Delta(H)\times\mathbb{R}^H \to \mathbb{R}^H \times \mathbb{R}^H_{\geq 0}$, $C,\mathbf{b}$ *s.t.*

$$s(\hat{Q},\hat{\boldsymbol{\mu}},Q,\boldsymbol{\mu}) = f(\hat{Q},\hat{\boldsymbol{\mu}})(Q,\boldsymbol{\mu}) - \int_{f_{|[\mathbf{b},(\hat{Q},\hat{\boldsymbol{\mu}})]}} f^{-1}(\mathbf{q})d\mathbf{q} + C. \tag{2}$$

Note that this is a path integral again. Since $f_{|[\mathbf{b},(\hat{Q},\hat{\boldsymbol{\mu}})]}$, the restriction of $f$ to the line $[\mathbf{b},(\hat{Q},\hat{\boldsymbol{\mu}})]$ 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. In Corollary A.16, we show the path independence of this extension of path integrals.

We now give an intuitive interpretation of Theorem 5.4, which we put in terms of *continuously* selling A-D shares and securities. For simplicity, we here imagine that $f$ is injective and continuous. The principal starts out giving the expert a quantity vector $f(\mathbf{b}) \in \mathbb{R}^H \times \mathbb{R}^H_{\geq 0}$ of A-D securities and shares for some constant (potentially negative) price $C$. (In our model, the expert has no choice of whether to accept this offer.) The expert can then move this quantity (through $\mathrm{im}(f)$), i.e., can buy or sell these assets. The prices of both buying and selling infinitesimal A-D shares and securities are $f^{-1}(\mathbf{q}) \in \mathbb{R}^H \times \mathbb{R}^H$, where $\mathbf{q}$ denotes the current quantity vector. Here, $f^{-1}$ can be any cyclically monotone increasing function. (By Lemma A.12, $f$ is cyclically monotone increasing if and only if $f^{-1}$ is cyclically monotone increasing.) Intuitively, this means that the more assets the expert has, the higher the marginal price of the assets. Further, by the path independence property, the $f^{-1}$ pricing is such that if the expert owns quantity vector $\mathbf{q}$, then after any sequence of infinitesimal purchases that end up in $\mathbf{q}$, the overall price paid by the expert for the sequence is zero.

Under this interpretation, how can we understand that the DSRs as described in Theorem 5.5 are proper? In the above interpretation, the expert buys and sells quantities. At what quantity vector $\mathbf{q}^*$ does he stop trading? Clearly, this must be one where the marginal price of the shares and securities is exactly equal to their value. That is, the quantity vector at which the expert stops trading must have the property $f^{-1}(\mathbf{q}^*) = (Q, \boldsymbol{\mu})$, where $(Q, \boldsymbol{\mu})$ is the expert's belief (given his recommendation). Thus, the value of the shares can be interpreted from the quantity bought. Of course, if we return to the DSR formalism, the expert does not report $\mathbf{q}^*$ (or move through the space of possible quantities, making infinitesimal shares). Instead, we can interpret the expert as directly reporting the price at which they want to cease trading. This is because the quantity resulting from a report of $(\hat{Q}, \hat{\boldsymbol{\mu}})$ is $f(\hat{Q}, \hat{\boldsymbol{\mu}})$, where the marginal price is $f^{-1}(f(\hat{Q}, \hat{\boldsymbol{\mu}})) = (\hat{Q}, \hat{\boldsymbol{\mu}})$ (assuming injectivity of $f$ for simplicity). As noted above, the expert prefers to stop trading at $(Q, \boldsymbol{\mu})$ – hence, the expert prefers honest reporting.

# 6 Multiple experts

Besides eliciting from a single expert, we are interested in designing mechanisms for eliciting information for decision making from multiple experts, ideally analogous 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. Based on that we will offer some speculation on what a decision making equivalent to prediction markets might look like.

## 6.1 Information structure in the multi-expert case

Again, we consider a principal who selects from a set of actions $A$. After she has taken an action, an outcome from $\Omega$ 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. We address this with 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 $H := \bigtimes_{i=1,...,n} H_i$. 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 by the experts and action $a$ is taken by the principal.

## 6.2 Truthful mechanisms

In general, a mechanism for a given private information structure is a special type of game $\Gamma$ 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 $\Gamma$. 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 $\arg\max_{a \in A} \mathbb{E}_P [u(O) \mid \mathbf{e}, a]$, is selected for any possible vector of observations on the expert's side. 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 (similar to the function that, as part of the game, selects an action) 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 $\sigma_i$ 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 specific generic truthful mechanisms. Here is one example of such a setup:

**Example 5.** 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 $\alpha_i$ 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 6.** 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 $\omega$. 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.

### 6.3 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 $(P(\cdot \mid \mathbf{e}_{-i}, a))_{a \in A, \mathbf{e}_{-i} \in H_{-i}}$.

**Theorem 6.1.** *Let $\Gamma$ 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:*

$$\Delta(H_{-i}) \times \Delta(\Omega)^{\Delta(H_{-i}) \times A} \times H_{-i} \times \Omega \to \mathbb{R} \colon (\hat{Q}_i, \hat{P}_i, e_{-i}, \omega) \mapsto u_i(\boldsymbol{\sigma}, (\mathbf{e}_{-i}, (\hat{Q}_i, \hat{P}_i)), \omega). \quad (3)$$

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 6.1 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 6. 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 6 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 6.1 operates on a useful model.

### 6.4 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 important, 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 expected score (assuming 0.1 represents his true beliefs over the outcome of the random event). We might want an elicitation mechanism for decision making – perhaps a kind of decision market (see Section 7.1) – 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. Generally, let $e_i^1, e_i^2 \in H_i$ be two possible pieces of evidence for expert $i$ that imply the same conditional A-D means $\boldsymbol{\mu} \in \mathbb{R}^{H_{-i}}$ and $Q \in \Delta(H_{-i})$. Then the expert must receive the same expected score from honestly reporting $e_i^1$ and honestly reporting $e_i^2$. This is the case even if the impact of these two pieces of information on the aggregate expert report differs substantially. For example, if for each $e_{-i}$, the best action given $e_{-i}$ is the same as the best action given $(e_i^1, e_{-i})$, then the report of $e_i^1$ has little impact.

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.

# 7 Related work

As noted in the introduction, our work is inspired by the literature on proper scoring rules and prediction markets. Further, we have noted in Section 5 (as well as the appendix) how results from this literature and from convex analysis can be used to derive our characterizations. In this section, we discuss further related work.

## 7.1 Conditional prediction markets

The earliest line of work that shares one of the goals of this paper (designing prediction market-like mechanisms for eliciting information for decision making) proposes *conditional prediction markets* (Hanson, 1999, 2002; Berg and Rietz, 2003; Hanson, 2006, 2013, Section IV), which work as follows. For each action $a \in A$, the principal sets up a prediction market dedicated to predicting the outcome conditional on $a$ being taken. At some point the principal takes the action that is best according to the market probability distributions. An outcome is observed and the prediction market for $a$ is resolved in the usual way. The prediction markets corresponding to all other actions are not resolved at all, i.e., no profits or losses are made on these markets. (If the market uses Arrow-Debreu securities, then all trades must be reversed.) As far as we are aware, 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). Compare our discussion of Chen et al.'s (2011; 2014) work below, who show how a variant of conditional prediction markets is indeed truthful.

## 7.2 Othman and Sandholm (2010)

As far as we can tell, Othman and Sandholm (2010) are the first to consider proper decision scoring rules as defined in our paper. In particular, 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. (The results of our Sections 3 and 4 are therefore trivial in the two-outcome case.) In Section 2.3.2, they give a characterization of *differentiable* proper decision scoring rules. A generalization to differentiable scoring rules $s$ for arbitrary $H, \Omega$ is given in Appendix E. The characterization's shape does not correspond straightforwardly to any of our general characterizations.

## 7.3 Chen, Kash, et al. (2014)

Chen, Kash, et al. (2014) also characterize scoring rules for decision making. (An alternative proof of their main result is also given by Frongillo and Kash (2014, Section E.1).) They consider the case $|H| = 1$ (though their results would be impacted little if we allowed the principal to have private evidence). Their key positive idea is the following. The expert reports an outcome prediction (i.e., a distribution over $\Omega$) for *each* action $a \in A$. Based on these predictions, the principal chooses an action $a$ randomly according to some distribution $\phi \in \Delta(A)$. For example, $\phi$ could assign $1-\epsilon$ probability to an action that is optimal according to the expert's reports and distribute the remaining probability $\epsilon$ equally among the other actions. Importantly, (if we want to strictly incentivize honest reports from the experts) $\phi$ must have full support, i.e., each action must be taken with positive probability. The expert is then scored for his outcome prediction for $a$ according to, say, Brier's scoring rule (or any other proper scoring rule for prediction, as characterized by Gneiting and Raftery, 2007). However, the score is divided by $\phi(a)$. Thus for each action $a$, the prediction is scored only with probability $\phi(a)$ but scaled up by $1/\phi(a)$. These cancel out in the expert's expected score term. Therefore, the expected scoring of the outcome prediction for $a$ is as though the prediction for $a$ was tested and scored according to Brier's scoring rule with probability 1. In particular, the expert is strictly incentivized to report honestly. Chen, Kash, et al. (2014)'s provide a general characterization of truthful mechanisms when the principal randomizes. In particular, they show that scaling up rewards by a factor of $1/\phi(a)$ is necessary.

Relative to our approach, Chen, Kash, et al.'s has at least two advantages. For one, it allows us to elicit full distributions for all actions rather than merely the expected utility of a single recommended action. Second, it allows us to construct decision markets that closely match the design of prediction markets (of either the market scoring rule or the Arrow-Debreu securities type) (cf. Wang and Pfeiffer, 2021). Our main concern with their approach is that randomization (and the required scaling in proportion to $1/\phi(a)$) comes with a number of drawbacks. We discuss these in detail in Section 2.

Chen, Kash, et al. (2014, Section 5) do also consider a setting similar to ours in which the expert recommends a single (optimal) action. 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.

## References


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.

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.

Boutilier, Craig (2012). "Eliciting Forecasts from Self-interested Experts: Scoring Rules for Decision Makers." In: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*. Ed. by Conitzer et al.

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

Chen, Yiling, Ian A. Kash, 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 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.

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.

Rochet, Jean (1987). “A necessary and sufficient condition for rationalizability in a quasi-linear context.” In: *Journal of Mathematical Economics* 16.2, pp. 191–200.

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.

Wang, Wenlong and Thomas Pfeiffer (2021). *Securities Based Decision Markets.* URL: https://arxiv.org/abs/2103.10011.

# 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, all of the below is known in the literature on cyclic monotonicity and convex functions.

## 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 $\mathbf{x}_0, \mathbf{x} \in \mathbb{R}^n$,

$$F(\mathbf{x}) \geq F(\mathbf{x}_0) + f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0). \tag{4}$$

**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

$$\begin{aligned} F(p\mathbf{x} + (1-p)\mathbf{y}) &= F(\mathbf{z}) + f(\mathbf{z})(p\mathbf{x} + (1-p)\mathbf{y} - \mathbf{z}) \\ &= p(F(\mathbf{z}) + f(\mathbf{z})(\mathbf{x} - \mathbf{z})) + (1-p)(F(\mathbf{z}) + f(\mathbf{z})(\mathbf{y} - \mathbf{z})) \\ &\underset{\text{Ineq. 4}}{\leq} pF(x) + (1-p)F(y). \end{aligned}$$

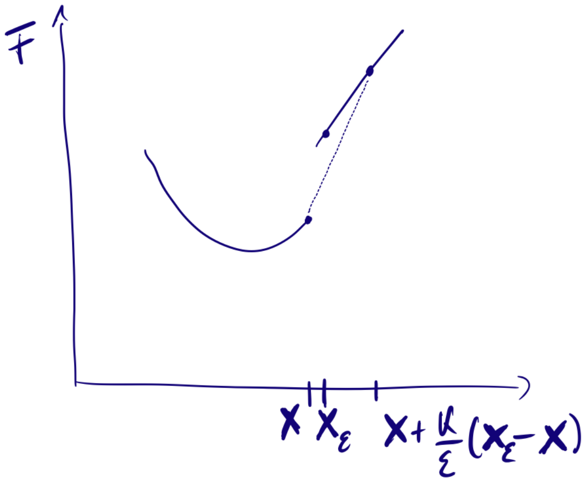

Figure 1: An illustration of the proof of Lemma A.4.

□

**Lemma A.2.** *Let $M$ be any index set. Let $F_i \colon U \to \mathbb{R}$ be convex functions for all $i \in M$. Then if the pointwise maximum function*

$$F^* \colon U \to \mathbb{R} \colon \mathbf{x} \mapsto \max_{i \in M} F_i(\mathbf{x})$$

*exists, it is convex.*

*Proof.* Let $\mathbf{x}_1, \mathbf{x}_2 \in U$, $p \in [0,1]$. Let $i^*$ be s.t. $F(p\mathbf{x}_1 + (1-p)\mathbf{x}_2) = F_{i^*}(p\mathbf{x}_1 + (1-p)\mathbf{x}_2)$. It is

$$\begin{aligned} & pF^*(\mathbf{x}_1) + (1-p)F^*(\mathbf{x}_2) \\ \geq\ & pF_{i^*}(\mathbf{x}_1) + (1-p)F_{i^*}(\mathbf{x}_2) \\ \geq\ & F_{i^*}(p\mathbf{x}_1 + (1-p)\mathbf{x}_2) \\ =\ & F^*(p\mathbf{x}_1 + (1-p)\mathbf{x}_2) \end{aligned}$$

□

**Lemma A.3.** *Let $\mathbf{x_1}, ..., \mathbf{x}_k \in \mathbb{R}^n$ and*

$$F \colon \left\{ \sum_{i=1}^{k} p_i \mathbf{x}_i \;\middle|\; \sum_{i=1}^{k} p_i = 1, \forall i : p_i \geq 0 \right\} \to \mathbb{R} \tag{5}$$

*be convex. Then $F(\mathbf{x}) \leq \max_{i=1,..,k} F(\mathbf{x}_i)$ for all $\mathbf{x}$ in $F$'s domain.*

**Lemma A.4.** *Let $U \subseteq \mathbb{R}^n$ be open and convex and $F \colon U \to \mathbb{R}$ be convex. Then $F$ is continuous.*

*Proof.* We prove the contrapositive. Let there be a discontinuity at $\mathbf{x}$. This means that there is $\delta > 0$ s.t. there are arbitrarily small $\epsilon > 0$ with $\mathbf{x}_\epsilon \in U$ with distance exactly $\epsilon$ from $\mathbf{x}$ s.t. $|F(\mathbf{x}_\epsilon) - F(\mathbf{x})| > \delta$. It follows that (at least) one of the following two must hold.

(1) There is $\delta > 0$ s.t. there are arbitrarily small $\epsilon > 0$ with $\mathbf{x}_\epsilon \in U$ with distance exactly $\epsilon$ from $\mathbf{x}$ s.t. $F(\mathbf{x}_\epsilon) - F(\mathbf{x}) > \delta$.

(2) There is $\delta > 0$ s.t. there are arbitrarily small $\epsilon > 0$ with $\mathbf{x}_\epsilon \in U$ with distance exactly $\epsilon$ from $\mathbf{x}$ s.t. $F(\mathbf{x}) - F(\mathbf{x}_\epsilon) > \delta$.

We first prove convexity in case 1. First, draw a convex polygon around $\mathbf{x}$ that contains some $\epsilon$ ball around $\mathbf{x}$. Because $U$ is open, if we make the polygon small enough, it lies fully in $U$. By Lemma A.3, $F$ is upper-bounded by some $L \in \mathbb{R}$.

Next, there is $K > 0$ s.t. for all $\epsilon > 0$,

$$\mathbf{x} + \frac{K}{\epsilon}(\mathbf{x}_\epsilon - \mathbf{x}) \tag{6}$$

is in the convex polygon. Intuitively, this is a point that arises from going in the direction of $\mathbf{x}_\epsilon$ for a distance $K$ that is constant and in particular independent of $\epsilon$. In particular, for small $\epsilon$ we walk *past* $\mathbf{x}_\epsilon$. It is,

$$\begin{aligned}
& \left(1 - \frac{\epsilon}{K}\right) F(\mathbf{x}) + \frac{\epsilon}{K} F\left(\mathbf{x} + \frac{K}{\epsilon}(\mathbf{x}_\epsilon - \mathbf{x}\right) \\
\leq \quad & \left(1 - \frac{\epsilon}{K}\right) F(\mathbf{x}) + \frac{\epsilon}{K} L \\
\overset{\epsilon \to 0}{\to} \quad & F(\mathbf{x}) \\
< \quad & F(\mathbf{x}_\epsilon) - \delta \\
= \quad & F\left(\left(1 - \frac{\epsilon}{K}\right)\mathbf{x} + \frac{\epsilon}{K}\left(\mathbf{x} + \frac{K}{\epsilon}(\mathbf{x}_\epsilon - \mathbf{x})\right)\right).
\end{aligned}$$

For small enough $\epsilon$, we thus find convexity violated.

Case 2 is mostly analogous. The main difference is that we need to consider

$$\mathbf{x}_\epsilon + \frac{K}{\epsilon}(\mathbf{x} - \mathbf{x}_\epsilon). \tag{7}$$

Thus, we now walk from $\mathbf{x}_\epsilon$ in the direction of (and for small $\epsilon$ past) $\mathbf{x}$. This way, we again walk from the smaller value of $F$ to the larger one. It then is

$$\begin{aligned}
& \left(1 - \frac{\epsilon}{K}\right) F(\mathbf{x}_\epsilon) + \frac{\epsilon}{K} F\left(\mathbf{x}_\epsilon + \frac{K}{\epsilon}(\mathbf{x} - \mathbf{x}_\epsilon)\right) \\
\leq \quad & \left(1 - \frac{\epsilon}{K}\right) F(\mathbf{x}_\epsilon) + \frac{\epsilon}{K} L \\
\overset{\epsilon \to 0}{\to} \quad & F(\mathbf{x}_\epsilon) \\
< \quad & F(\mathbf{x}) - \delta \\
= \quad & F\left(\left(1 - \frac{\epsilon}{K}\right)\mathbf{x}_\epsilon + \frac{\epsilon}{K}\left(\mathbf{x}_\epsilon + \frac{K}{\epsilon}(\mathbf{x} - \mathbf{x}_\epsilon)\right)\right).
\end{aligned}$$

Again we find convexity violated. Here the $\overset{\epsilon \to 0}{\to}$ line should be interpreted as the difference of the left and right-hand side going to zero. □

**Lemma A.5.** *Let $F$ be differentiable with gradient $\nabla F$. Further let $F$ have a subgradient $f$. Then $f = \nabla F$.*

*Proof.* We prove this by proving equality of each of the entries of $f, \nabla F$. The $k$-th entry of $\nabla F(\mathbf{x})$ is simply $\frac{d}{dx_k}F(\mathbf{x})$, the derivative w.r.t. $x_k$. By definition of differentiability,

$$\lim_{h\to 0} \frac{f(\mathbf{x}+he_k)-f(\mathbf{x})}{h} = \frac{d}{dx_k}F(\mathbf{x})$$

for all $\mathbf{x}$, where $e_k$ is the vector whose $k$-th entry is 1 and whose other entries are 0. Equivalently,

$$\lim_{h\downarrow 0} \frac{f(\mathbf{x}+he_k)-f(\mathbf{x})}{h} = \frac{d}{dx_k}F(\mathbf{x}) = \lim_{h\downarrow 0} \frac{f(\mathbf{x})-f(\mathbf{x}-he_k)}{h}, \tag{8}$$

where both limits let $h$ go to zero from above.

For $f$ to be a subgradient of $F$, it must be for all $h \geq 0$,

$$F(\mathbf{x}+he_k) \geq F(\mathbf{x}) + f(\mathbf{x})(he_k)$$

as well as

$$F(\mathbf{x}-he_k) \geq F(\mathbf{x}) - f(\mathbf{x})(he_k).$$

Solving both inequalities for $f(\mathbf{x})$, we obtain that for all $h > 0$,

$$\frac{f(\mathbf{x})-f(\mathbf{x}-he_k)}{h} \leq f'(\mathbf{x})e_k \leq \frac{f(\mathbf{x}+he_k)-f(\mathbf{x})}{h}.$$

From equation 8, it follows that $f(\mathbf{x})e_k = \frac{d}{dx_k}F(\mathbf{x})$ as claimed. □

## A.2 Convex functions are path integrals of their subgradients

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

**Theorem A.6.** *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 $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}$.,*

$$\int_{\mathbf{x}_1}^{\mathbf{x}_2} f(\mathbf{z})d\mathbf{z} = F(\mathbf{x}_2) - F(\mathbf{x}_1), \tag{9}$$

*where the path integral is path-independent.*

This result is analogous to the fundamental theorem of calculus for line integrals (a.k.a. the gradient theorem). Rockafellar (1970, Corollary 24.2.1) first gave the $n = 1$ special case of this results (which in can be seen as a version of the fundamental theorem of calculus). Theorem A.6 follows directly from Rockafellar's result (cf. Frongillo and Kash, 2014, Appendix A, Fact 3). We omit a direct proof of Theorem A.6 here, because it can be proven using a subset of the ideas in the proof of one of the directions of Lemma D.3.

## 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 $\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_k \in M$ with $k \geq 1$ and $\mathbf{x}_0 = \mathbf{x}_k$,

$$\sum_{i=1}^{k} f(\mathbf{x}_{i-1})(\mathbf{x}_i - \mathbf{x}_{i-1}) \leq 0. \tag{10}$$

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

**Lemma A.7.** *A function $f\colon M \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is cyclically monotone increasing if and only if for all $\mathbf{z}_0, ..., \mathbf{z}_k \in M$ with $\mathbf{z}_0 = \mathbf{z}_k$,*

$$\sum_{i=1}^{k} f(\mathbf{z}_i)(\mathbf{z}_i - \mathbf{z}_{i-1}) \geq 0. \tag{11}$$

*Proof.* Multiplying Ineq. 10 by $-1$ yields

$$f(\mathbf{x}_0)(\mathbf{x}_0 - \mathbf{x}_1) + f(\mathbf{x}_1)(\mathbf{x}_1 - \mathbf{x}_2) + ... + f(\mathbf{x}_{n-1})(\mathbf{x}_{n-1} - \mathbf{x}_n) + f(\mathbf{x}_n)(\mathbf{x}_n - \mathbf{x}_0) \geq 0. \tag{12}$$

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

$$f(\mathbf{z}_n)(\mathbf{z}_n - \mathbf{z}_{n-1}) + f(\mathbf{z}_{n-1})(\mathbf{z}_{n-1} - \mathbf{z}_{n-2}) + ... + f(\mathbf{z}_1)(\mathbf{z}_1 - \mathbf{z}_0) + f(\mathbf{z}_0)(\mathbf{z}_0 - \mathbf{z}_n) \geq 0. \tag{13}$$

And this is the same as Ineq. 11, except that the order of the first $n$ summands is reversed. □

**Lemma A.8.** *If $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is cyclically monotone increasing, then for any $\mathbf{x}_a, \mathbf{x}_b$, the function $[0,1] \to \mathbb{R}^n \colon t \to (\mathbf{x}_b - \mathbf{x}_a) f(\mathbf{x}_a + t(\mathbf{x}_b - \mathbf{x}_a))$ is monotone increasing.*

Conditions equivalent to this monotonicity on the line are sometimes used for defining monotonicity (a.k.a. two-cycle monotonicity) for functions $\mathbb{R}^n \to \mathbb{R}^n$. Note that the lemma only holds in one direction.

*Proof.* Let $t_1, t_2 \in [0,1]$, $t_1 < t_2$. Define $\mathbf{x}_i = \mathbf{x}_a + t(\mathbf{x}_b - \mathbf{x}_a$ for $i = 1, 2$. By definition of cyclic monotonicity and Lemma A.7,

$$f(\mathbf{x}_2)(\mathbf{x}_2 - \mathbf{x}_1) + f(\mathbf{x}_1)(\mathbf{x}_1 - \mathbf{x}_2) \geq 0.$$

By simple arithmetics, this is equivalent to

$$f(\mathbf{x}_2)(\mathbf{x}_2 - \mathbf{x}_1) \geq f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1).$$

Inserting the definitions of $\mathbf{x}_1, \mathbf{x}_2$ and dividing by $t$ yields that

$$(\mathbf{x}_b - \mathbf{x}_a) f(\mathbf{x}_a + t_2(\mathbf{x}_b - \mathbf{x}_a)) \geq (\mathbf{x}_b - \mathbf{x}_a) f(\mathbf{x}_a + t_1(\mathbf{x}_b - \mathbf{x}_a))$$

as claimed. □

**Lemma A.9.** *Let $f\colon \mathbb{R} \to \mathbb{R}$ be a function on the reals. Then $f$ is cyclically monotone increasing if and only if $f$ is monotone increasing.*

**Lemma A.10.** *Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be cyclically monotone increasing. Then $f$ is path-independently integrable.*

*Proof.* First note that the line integral of $f$ on any $[\mathbf{x}_a, \mathbf{x}_b]$ is by definition a regular integral of the real function in Lemma A.8 on the interval $[0,1]$. Because by Lemma A.8 this function is monotone, it is integrable (e.g., Rudin 1976, Theorem 6.9). Hence, $f$ is (path-)integrable on $[\mathbf{x}_a, \mathbf{x}_b]$.

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

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

This can be seen as follows. Because $f$ is cyclically monotone increasing, it is for all $\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_k \in \mathbb{R}^n$ with $\mathbf{x}_0 = \mathbf{x}_k$ by definition and Lemma A.7,

$$\sum_{i=1}^{k} f(\mathbf{x}_i)(\mathbf{x}_i - \mathbf{x}_{i-1}) \geq 0 \geq \sum_{i=1}^{k} f(\mathbf{x}_{i-1})(\mathbf{x}_i - \mathbf{x}_{i-1}). \tag{14}$$

Now, because $f$ is integrable, if we let the cycle $\mathbf{x}_1, ..., \mathbf{x}_k$ become arbitrarily fine approximations of $\gamma$, the left and right sum converge to the integral along $\gamma$ and therefore to the same value. By Ineq. 14, that value must be 0. □

**Lemma A.11.** *If $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is strictly two-cycle monotone, then $f$ is injective.*

*Proof.* By contradiction. Assume there are $\mathbf{x}_a, \mathbf{x}_b \in \mathbb{R}^n$ with $\mathbf{x}_a \neq \mathbf{x}_b$ but $f(\mathbf{x}_a) = f(\mathbf{x}_b)$. Then

$$\mathbf{x}_a(f(\mathbf{x}_a) - f(\mathbf{x}_b)) + \mathbf{x}_b(f(\mathbf{x}_b) - f(\mathbf{x}_a)) = 0$$

in contradiction to strict two-cycle monotonicity of $f$. □

**Lemma A.12.** *Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be strictly cyclically monotone. Then $f^{-1}\colon \mathrm{im}(f) \to \mathbb{R}^n$, which exists as per the above, is strictly cyclically monotone.*

*Proof.* Let $\mathbf{y}_0, \mathbf{y}_1, ..., \mathbf{y}_k \in \mathrm{im}(f)$ with $\mathbf{y}_0 = \mathbf{y}_k$. Further, let $\mathbf{x}_i = f^{-1}(\mathbf{y}_i)$. Then

$$\begin{aligned} & \sum_{i=1}^{k} \mathbf{y}_i(f^{-1}(\mathbf{y}_i) - f^{-1}(\mathbf{y}_{i-1})) \\ = \quad & \sum_{i=1}^{k} f(\mathbf{x}_i)(\mathbf{x}_i - \mathbf{x}_{i-1}) \\ \underset{\text{Lemma A.7}}{\geq} \quad & 0 \end{aligned}$$

□

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

**Theorem A.13** (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.*

We omit proofs here, because the two directions are analogous to other proofs we give below. If $f$ is a subgradient, then it is easy to show that $f$ is cyclically monotone increasing, compare the proof of Lemma D.2. The other direction is harder and analogous to the proof of one of the two directions of Lemma D.3.

## 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 $\gamma$ that are cyclically monotone increasing, but not necessarily continuous.

To make this well-defined we first extend $\gamma$ to a set-valued function $\bar{\gamma}$ such that $\bar{\gamma}([\mathbf{a}, \mathbf{b}])$ is a curve. So let $\gamma$ 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 $\gamma$ jumps immediately after $\mathbf{x}$. Imagine further that $\gamma$ 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 $\gamma$ jumps to the left or on both sides of $\mathbf{x}$, we define $\bar{\gamma}(\mathbf{x})$ analogously. If $\gamma$ 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

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

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.14.** *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.*

This follows directly from the monotonicity on the line as per Lemma A.8.

**Theorem A.15.** *Let $f \colon \mathbb{R}^n \to \mathbb{R}^n$ be cyclically monotone increasing and $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$. Then*

$$\int_{f_{|[\mathbf{a},\mathbf{b}]}} f^{-1}_{|[\mathbf{a},\mathbf{b}]}(\mathbf{y}) d\mathbf{y} = \mathbf{b} f(\mathbf{b}) - \mathbf{a} f(\mathbf{a}) - \int_{\mathbf{a}}^{\mathbf{b}} f(\mathbf{x}) d\mathbf{x}. \tag{15}$$

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* (cyclically) monotone and discontinuous. The second is that we allow $n > 1$, i.e., that we generalize the result to path integrals.

As long as we keep $n = 1$, the result remains intuitive and we give the typical type of illustration in Figure 2. As usual, the integral $\int_a^b f(x)dx$ is the area under the curve of $f$ between $a$ and $b$. Similarly, the integral $\int_{f(a)}^{f(b)} f^{-1}(y)dy$, with our technical extension, is the light blue area, which is the area under the curve between $f(a)$ and $f(b)$ if we turn the graph 90 degrees counter-clockwise and fill the discontinuity with a straight line. By inspection, we can see that the square with area $f(b)b$ can be partitioned into a square with area $f(a)a$ and the area under curves described by the integrals $\int_a^b f(x)dx$ and $\int_{f(a)}^{f(b)} f^{-1}(y)dy$. This is exactly the claim of the theorem in case $n = 1$.

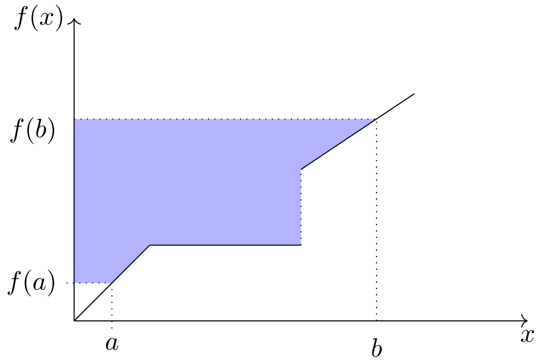


Figure 2: An illustration of Theorem A.15 for $n = 1$.

*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.14.

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. □

**Corollary A.16.** *Let $f \colon \mathbb{R}^n \to \mathbb{R}^n$ be cyclically monotone increasing. Then the path integral of $f^{-1}$ is path-independent. That is, for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$,*

$$\int_{f_{|[\mathbf{a},\mathbf{b}]}} f^{-1}_{|[\mathbf{a},\mathbf{b}]}(\mathbf{y}) d\mathbf{y} + \int_{f_{|[\mathbf{b},\mathbf{c}]}} f^{-1}_{|[\mathbf{b},\mathbf{c}]}(\mathbf{y}) d\mathbf{y} = \int_{f_{|[\mathbf{a},\mathbf{c}]}} f^{-1}_{|[\mathbf{a},\mathbf{c}]}(\mathbf{y}) d\mathbf{y}. \tag{16}$$

*Proof.*

$$\begin{aligned}
&\int_{f_{|[\mathbf{a},\mathbf{b}]}} f^{-1}_{|[\mathbf{a},\mathbf{b}]}(\mathbf{y})d\mathbf{y} + \int_{f_{|[\mathbf{b},\mathbf{c}]}} f^{-1}_{|[\mathbf{b},\mathbf{c}]}(\mathbf{y})d\mathbf{y} \\
\underset{\text{Theorem A.15}}{=} &\left(\mathbf{b}f(\mathbf{b}) - \mathbf{a}f(\mathbf{a}) - \int_{\mathbf{a}}^{\mathbf{b}} f(\mathbf{x})d\mathbf{x}\right) + \left(\mathbf{c}f(\mathbf{c}) - \mathbf{b}f(\mathbf{b}) - \int_{\mathbf{b}}^{\mathbf{c}} f(\mathbf{x})d\mathbf{x}\right) \\
= &\ \mathbf{c}f(\mathbf{c}) - \mathbf{a}f(\mathbf{a}) - \left(\int_{\mathbf{a}}^{\mathbf{b}} f(\mathbf{x})d\mathbf{x} + \int_{\mathbf{b}}^{\mathbf{c}} f(\mathbf{x})d\mathbf{x}\right) \\
\underset{\text{Lemma A.10}}{=} &\ \mathbf{c}f(\mathbf{c}) - \mathbf{a}f(\mathbf{a}) - \int_{\mathbf{a}}^{\mathbf{c}} f(\mathbf{x})d\mathbf{x} \\
\underset{\text{Theorem A.15}}{=} &\int_{f_{|[\mathbf{a},\mathbf{c}]}} f^{-1}_{|[\mathbf{a},\mathbf{c}]}(\mathbf{y})d\mathbf{y}.
\end{aligned}$$

□

# B Proofs for Section 3

## B.1 Expected proper scoring under honest reporting is convex

We here prove an easy lemma that we will use our proofs. This type of result/proof is well-known in the literature on proper scoring rules for prediction.

**Lemma B.1.** *If $s$ is proper, then the function $(Q, P_\alpha) \mapsto \mathbb{E}_{E\sim Q, O\sim P_\alpha}[s(Q, P_\alpha, E, O)]$ is convex.*

*Proof.* Because $s$ is proper,

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

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 $Q, P_\alpha$. 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 and therefore convex functions. By Lemma A.2, the point-wise maximum of affine convex functions is convex. □

## B.2 Continuity under honest reporting

**Lemma B.2.** *If $s$ is proper, then $\mathbb{E}_{E\sim Q, O\sim P_\alpha}[s(Q, P_\alpha, E, O)]$ is continuous in $P_\alpha, Q$ in the set of $P_\alpha, Q$ with full support.*

*Proof of Lemma B.2 with convex analysis.* By Lemma B.1, the given function is convex. By Lemma A.4, it is therefore continuous on the given (open) set. □

*Proof of Lemma B.2 without convex analysis.* We prove the contrapositive. Let there be a discontinuity at $(Q, P_\alpha)$. This means there is $\delta > 0$ s.t. for all $\epsilon > 0$ there is $(Q_\epsilon, P_{\alpha,\epsilon})$ s.t.

$$|\mathbb{E}_{E\sim Q, O\sim P_\alpha}[s(Q, P_\alpha, E, O)] - \mathbb{E}_{E\sim Q_\epsilon, O\sim P_{\alpha,\epsilon}}[s(Q_\epsilon, P_{\alpha,\epsilon}, E, O)]| > \delta.$$

It follows that for the given $\delta$ (at least) one of the following two must hold:

(1) For all $\epsilon > 0$ there is $(Q_\epsilon, P_{\alpha,\epsilon})$ s.t.

$$\mathbb{E}_{E\sim Q, O\sim P_\alpha}\left[s(Q, P_\alpha, E, O)\right] - \mathbb{E}_{E\sim Q_\epsilon, O\sim P_{\alpha,\epsilon}}\left[s(Q_\epsilon, P_{\alpha,\epsilon}, E, O)\right] > \delta.$$

(2) For all $\epsilon > 0$ there is $(Q_\epsilon, P_{\alpha,\epsilon})$ s.t.

$$\mathbb{E}_{E\sim Q, O\sim P_\alpha}\left[s(Q, P_\alpha, E, O)\right] - \mathbb{E}_{E\sim Q_\epsilon, O\sim P_{\alpha,\epsilon}}\left[s(Q_\epsilon, P_{\alpha,\epsilon}, E, O)\right] < -\delta.$$

Consider case 1. The idea is now that if the expert's belief is $(Q_\epsilon, P_{\alpha,\epsilon})$, he could report $(Q, P_\alpha)$ because intuitively speaking as $\epsilon$ approaches 0, the expected score cannot tell whether the true distribution is $(Q_\epsilon, P_{\alpha,\epsilon})$ or $(Q, P_\alpha)$:

$$\begin{aligned}\mathbb{E}_{E\sim Q_\epsilon, O\sim P_{\alpha,\epsilon}}\left[s(Q, P_\alpha, E, O)\right] &\overset{\epsilon\to 0}{\to} \mathbb{E}_{E\sim Q, O\sim P_\alpha}\left[s(Q, P_\alpha, E, O)\right] \\ &> \mathbb{E}_{E\sim Q_\epsilon, O\sim P_{\alpha,\epsilon}}\left[s(Q_\epsilon, P_{\alpha,\epsilon}, E, O)\right] + \delta\end{aligned}$$

The convergence is due to the fact that the expectation is continuous in the distribution(s) that it is over (in this case $Q_\epsilon, P_{\alpha,\epsilon}$). It follows that for small enough $\epsilon$,

$$\mathbb{E}_{E\sim Q_\epsilon, O\sim P_{\alpha,\epsilon}}\left[s(Q, P_\alpha, E, O)\right] > \mathbb{E}_{E\sim Q_\epsilon, O\sim P_{\alpha,\epsilon}}\left[s(Q_\epsilon, P_{\alpha,\epsilon}, E, O)\right].$$

By definition, this means that $s$ is not proper, as claimed.

Case 2 is dealt with analogously by reporting $(Q_\epsilon, P_{\alpha,\epsilon})$ when the true belief is $(Q, P_\alpha)$, for small enough $\epsilon$. □

## B.3 No expert preferences under honest reporting

**Lemma B.3.** *Let $s$ be a proper DSR and $P_\alpha, P'_{\alpha'} \in \Delta(\Omega)^H$ be s.t. for all $e \in H$*

$$\min_{\omega\in\Omega} u(\omega) < \mathbb{E}_{O\sim P_\alpha}\left[u(O) \mid e\right] = \mathbb{E}_{O\sim P'_{\alpha'}}\left[u(O) \mid e\right] < \max_{\omega\in\Omega} u(\omega). \tag{17}$$

*Then*

$$\mathbb{E}_{E\sim Q, O\sim P_\alpha}\left[s(Q, P_\alpha, E, O)\right] = \mathbb{E}_{E\sim Q, O\sim P_{\alpha'}}\left[s(Q, P_{\alpha'}, E, O)\right]. \tag{18}$$

*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 $e \in H$:

$$R_\mathbf{p}(\omega_H \mid e) = p_e \tag{19}$$

$$R_\mathbf{p}(\omega_L \mid e) = 1 - p_e \tag{20}$$

$$R_\mathbf{p}(\omega \mid e) = 0 \text{ for all } \omega \in \{\omega_L, \omega_H\} \tag{21}$$

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

$$\mathbb{E}_{E\sim Q, O\sim P_\alpha}\left[s(Q, P_\alpha, E, O)\right] \geq \mathbb{E}_{E\sim Q, O\sim R_\mathbf{p}}\left[s(Q, R_\mathbf{p}, E, O)\right] \tag{22}$$

whenever all the means of $R_\mathbf{p}$ are element-wise at most the means of $P_\alpha$. The reverse inequality has to hold if the means of $R_\mathbf{p}$ are at at least as high the means of $P_\alpha$. 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 B.2, it follows that

$$\mathbb{E}_{E\sim Q, O\sim P_\alpha}\left[s(Q, P_\alpha, E, O)\right] = \mathbb{E}_{E\sim Q, O\sim R_\mathbf{p}}\left[s(Q, R_\mathbf{p}, E, O)\right] \tag{23}$$

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

It is worth that Lemma B.3 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.

**Example 7.** 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 A-D share for every $e$ and an extra A-D share 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 $\alpha, P_\alpha$ 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 $P_\alpha$ versus $P_{\alpha'}$.

## B.4 Proof of Theorem 3.1

**Theorem 3.1.** *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 $e \in H$, $\min_{\omega\in\Omega} u(\omega) < \mathbb{E}_{O\sim P_\alpha}[u(O) \mid e] = \mathbb{E}_{O\sim\hat{P}_\alpha}[u(O) \mid e] < \max_{\omega\in\Omega} u(\omega)$ and $\mathrm{supp}(P_\alpha(\cdot \mid e)) \subseteq \mathrm{supp}(\hat{P}_\alpha(\cdot \mid e))$. Then $\mathbb{E}_{E\sim Q, O\sim P_\alpha}[s(Q, P_\alpha, E, O)] = \mathbb{E}_{E\sim Q, O\sim P_\alpha}\left[s(Q, \hat{P}_\alpha, E, O)\right]$.*

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

$$P'_\alpha = \frac{1}{1-p}\left(\hat{P}_\alpha - pP_\alpha\right). \tag{24}$$

Because $\mathrm{supp}(P_\alpha(\cdot \mid e)) \subseteq \mathrm{supp}(\hat{P}_\alpha(\cdot \mid e))$ 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

$$\hat{P}_\alpha = pP_\alpha + (1-p)P'_\alpha. \tag{25}$$

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

Then

$$\begin{aligned}
& \mathbb{E}_{\hat{P}_\alpha}\left[s(Q, \hat{P}_\alpha, E, O)\right] && (26)\\
= \;& p\mathbb{E}_{P_\alpha}\left[s(Q, \hat{P}_\alpha, E, O)\right] + (1-p)\mathbb{E}_{P'_\alpha}\left[s(Q, \hat{P}_\alpha, E, O)\right] && (27)\\
\underset{s\text{ is proper}}{\leq} \;& p\mathbb{E}_{P_\alpha}\left[s(Q, \hat{P}_\alpha, E, O)\right] + (1-p)\mathbb{E}_{P'_\alpha}\left[s(Q, P'_\alpha, E, O)\right] && (28)\\
\underset{s\text{ is proper}}{\leq} \;& p\mathbb{E}_{P_\alpha}\left[s(Q, P_\alpha, E, O)\right] + (1-p)\mathbb{E}_{P'_\alpha}\left[s(Q, P'_\alpha, E, O)\right] && (29)\\
\underset{\text{Lemma } B.3}{=} \;& \mathbb{E}_{\hat{P}_\alpha}\left[s(Q, \hat{P}_\alpha, E, O)\right]. && (30)
\end{aligned}$$

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

## B.5 Proof of Lemma 3.2

**Lemma 3.2.** *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 $e \in H$, $\min_{\omega\in\Omega} u(\omega) < \mathbb{E}_{O\sim P_\alpha}[u(O) \mid e] < \max_{\omega\in\Omega} u(\omega)$. Further, let $\omega_1, \omega_2 \in \operatorname{supp}(\hat{P}_\alpha(\cdot \mid e))$ for some $e \in H$. Then $s(Q, P_\alpha, e, \omega_1) = s(Q, P_\alpha, e, \omega_2)$.*

*Proof.* If $u(\omega_{1/2}) = \mathbb{E}_{O\sim P_\alpha}[u(O) \mid e]$, the result follows (almost) immediately from Lemma 3.1. Else, there exists some $\omega_3 \in \operatorname{supp}(P_\alpha(\cdot \mid e))$ and $p \in (0, 1]$ s.t.

$$O_1 \mid e = p * \omega_1 + (1 - p) * \omega_3 \tag{31}$$
$$O_2 \mid e = p * \omega_2 + (1 - p) * \omega_3 \tag{32}$$

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

$$\begin{aligned}
& Q(e)ps(Q, P_\alpha, e, \omega_1) + Q(e)(1-p)s(Q, P_\alpha, e, \omega_3) + (1 - Q(e))\mathbb{E}[s(Q, P_\alpha, E, O) \mid E \neq e] \\
= \quad & Q(e)\mathbb{E}[s(Q, P_\alpha, e, O_1) \mid e] + (1 - Q(e))\mathbb{E}[s(Q, P_\alpha, E, O_1) \mid E \neq e'] \\
= \quad & \mathbb{E}[s(Q, P_\alpha, e, O_1)] \\
\underset{\text{Lemma 3.1}}{=} \quad & \mathbb{E}[s(Q, P_{O_1}, e, O_1)] \\
\underset{\text{Lemma } B.3}{=} \quad & \mathbb{E}[s(Q, P_{O_2}, e, O_2)] \\
\underset{\text{Lemma 3.1}}{=} \quad & \mathbb{E}[s(Q, P_\alpha, e, O_2)] \\
= \quad & Q(e)\mathbb{E}[s(Q, P_\alpha, e, O_2) \mid e] + (1 - Q(e))\mathbb{E}[s(Q, P_\alpha, E, O_2) \mid E \neq e] \\
= \quad & Q(e)ps(Q, P_\alpha, e, \omega_2) + Q(e)(1-p)s(Q, P_\alpha, e, \omega_3) + (1 - Q(e))\mathbb{E}[s(Q, P_\alpha, E, O) \mid E \neq e]
\end{aligned}$$

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 $f_Y \colon \Delta(H) \times \mathbb{R}^H \to \mathbb{R}^H_{\geq 0}, f_E \colon \Delta(H) \times \mathbb{R}^H \to \mathbb{R}^H$, $g \colon \Delta(H) \times \mathbb{R}^H \to \mathbb{R}$, s.t. for all $\hat{Q}, \hat{\boldsymbol{\mu}}, e, y$, $s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, y) = f_Y(\hat{Q}, \hat{\boldsymbol{\mu}})_e y + f_E(\hat{Q}, \hat{\boldsymbol{\mu}})_e + g(\hat{Q}, \hat{\boldsymbol{\mu}})$.*

*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

$$\mathbb{E}_X\left[s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, X)\right] = s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, x). \tag{33}$$

From this, the claimed 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

$$\begin{aligned}
& \hat{Q}(e)p\mathbb{E}\left[s(\hat{Q},\hat{\boldsymbol{\mu}},e,X)\right] + \hat{Q}(e)(1-p)s(\hat{Q},\hat{\boldsymbol{\mu}},e,x') + (1-\hat{Q}(e))\mathbb{E}\left[s(\hat{Q},\hat{\boldsymbol{\mu}},E,Y) \mid E \neq e\right] && (34)\\
= \; & \hat{Q}(e)\mathbb{E}\left[s(\hat{Q},\hat{\boldsymbol{\mu}},e,Y)\right] + (1-\hat{Q}(e))\mathbb{E}\left[s(\hat{Q},\hat{\boldsymbol{\mu}},E,Y) \mid E \neq e\right] && (35)\\
= \; & \mathbb{E}\left[s(\hat{Q},\hat{\boldsymbol{\mu}},e,Y)\right] && (36)\\
= \; & \mathbb{E}\left[s(\hat{Q},\hat{\boldsymbol{\mu}},e,\tilde{Y})\right] && (37)\\
= \; & \hat{Q}(e)\mathbb{E}\left[s(\hat{Q},\hat{\boldsymbol{\mu}},e,\tilde{Y})\right] + (1-\hat{Q}(e))\mathbb{E}\left[s(\hat{Q},\hat{\boldsymbol{\mu}},E,\tilde{Y}) \mid E \neq e\right] && (38)\\
= \; & \hat{Q}(e)p\mathbb{E}\left[s(\hat{Q},\hat{\boldsymbol{\mu}},e,x)\right] + \hat{Q}(e)(1-p)s(\hat{Q},\hat{\boldsymbol{\mu}},e,x') + (1-\hat{Q}(e))\mathbb{E}\left[s(\hat{Q},\hat{\boldsymbol{\mu}},E,\tilde{Y}) \mid E \neq e\right] && (39)
\end{aligned}$$

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 $f\colon \Delta(H_{-i}) \times \mathbb{R}^{H_{-i}} \to \mathbb{R}^{H_{-i}} \times \mathbb{R}_{\geq 0}^{H_{-i}}$ and $F\colon \Delta(H_{-i}) \times \mathbb{R}^{H_{-i}} \to \mathbb{R}$ s.t. $s(\hat{Q},\hat{\boldsymbol{\mu}},Q,\boldsymbol{\mu}) =$*

$f(\hat{Q}, \hat{\boldsymbol{\mu}})((Q, \boldsymbol{\mu}) - (\hat{Q}, \hat{\boldsymbol{\mu}})) + F(\hat{Q}, \hat{\boldsymbol{\mu}})$*, where* $F$ *is convex and* $f$ *is a subgradient of* $F$*. Further,* $s$ *is strictly proper w.r.t. the recommendation for* $e$ *if* $f_{Y,e} > 0$*.*

*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

$$\begin{aligned} s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) &= f(\hat{Q}, \hat{\boldsymbol{\mu}})((Q, \boldsymbol{\mu}) - (\hat{Q}, \hat{\boldsymbol{\mu}})) + F(\hat{Q}, \hat{\boldsymbol{\mu}}) & (40)\\ &\underset{\text{Ineq. 4}}{\leq} F(Q, \boldsymbol{\mu}) & (41)\\ &= F(Q, \boldsymbol{\mu}) + f(Q, \boldsymbol{\mu})((Q, \boldsymbol{\mu}) - (Q, \boldsymbol{\mu})) & (42)\\ &= s(Q, \boldsymbol{\mu}, Q, \boldsymbol{\mu}). & (43) \end{aligned}$$

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

$$s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) = f(\hat{Q}, \hat{\boldsymbol{\mu}})(Q, \boldsymbol{\mu}) + g(\hat{Q}, \hat{\boldsymbol{\mu}}) = f(\hat{Q}, \hat{\boldsymbol{\mu}})(\hat{Q}, \hat{\boldsymbol{\mu}}) + g(\hat{Q}, \hat{\boldsymbol{\mu}}) + f(\hat{Q}, \hat{\boldsymbol{\mu}})((Q, \hat{\boldsymbol{\mu}}) - (\hat{Q}, \hat{\boldsymbol{\mu}})). \quad (44)$$

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

$$F(Q, \boldsymbol{\mu}) = s(Q, \boldsymbol{\mu}, Q, \boldsymbol{\mu}) \underset{s \text{ proper}}{\geq} s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) = F(\hat{Q}, \hat{\boldsymbol{\mu}}) + f(\hat{Q}, \hat{\boldsymbol{\mu}})((Q, \boldsymbol{\mu}) - (\hat{Q}, \hat{\boldsymbol{\mu}})). \quad (45)$$

This is exactly the subgradient inequality (Ineq. 4). □

## 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.*

$$s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) = f(\hat{Q}, \hat{\boldsymbol{\mu}})((Q, \boldsymbol{\mu}) - (\hat{Q}, \hat{\boldsymbol{\mu}})) + \int_{\mathbf{b}}^{(\hat{Q}, \hat{\boldsymbol{\mu}})} f(\mathbf{z}) d\mathbf{z} + C. \quad (1)$$

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* $(Q_a, \boldsymbol{\mu}_a), (Q_b, \boldsymbol{\mu}_b)$*,*

$$(f(Q_a, \boldsymbol{\mu}_a) - f(Q_b, \boldsymbol{\mu}_b))(Q_b, \boldsymbol{\mu}_b) \leq g(Q_a, \boldsymbol{\mu}_a) - g(Q_b, \boldsymbol{\mu}_b) \leq (f(Q_a, \boldsymbol{\mu}_a) - f(Q_b, \boldsymbol{\mu}_b))(Q_a, \boldsymbol{\mu}_a). \quad (46)$$

*Further,* $s$ *is strictly proper if and only if Ineq. 46 is strict (in both directions) whenever* $\boldsymbol{\mu}_a \neq \boldsymbol{\mu}_b$*.*

Note that the two sides of the inequality are equivalent under renaming. Also notice the similarity (but not equivalence!) of the inequality to the subgradient inequality.

Here is an intuition for the result. For simplicity, assume $g(Q_a, \boldsymbol{\mu}_a) > g(Q_b, \boldsymbol{\mu}_b)$. Then $g(Q_a, \boldsymbol{\mu}_a) - g(Q_b, \boldsymbol{\mu}_b)$ is the extra money that the expert has to pay to obtain $f(Q_a, \boldsymbol{\mu}_a)$

instead of $f(Q_b, \boldsymbol{\mu}_b)$ shares and securities. For $s$ (as defined by $f, g$ to be (strictly) proper, this price increase must be (strictly) worth it if the true belief is $Q_a, \boldsymbol{\mu}_a$ (this is the right-hand inequality). Also, the price increase must be (strictly) not worth it, if the true belief is $Q_b, \boldsymbol{\mu}_b$.

*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

$$(f(Q_a, \boldsymbol{\mu}_a) - f(Q_b, \boldsymbol{\mu}_b))(Q_b, \boldsymbol{\mu}_b) \leq g(Q_a, \boldsymbol{\mu}_a) - g(Q_b, \boldsymbol{\mu}_b). \tag{47}$$

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. 46 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. 46 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*

$$g(\hat{Q}, \hat{\boldsymbol{\mu}}) = f(\hat{Q}, \hat{\boldsymbol{\mu}})(\hat{Q}, \hat{\boldsymbol{\mu}}) - \int_{\mathbf{b}}^{(\hat{Q}, \hat{\boldsymbol{\mu}})} f(\mathbf{x}) d\mathbf{x} + C, \tag{48}$$

*for any $C \in \mathbb{R}, \mathbf{b} \in \Delta(H) \times \mathbb{R}^H$ are exactly the functions that satisfy Ineq. 46 for all $(Q_a, \boldsymbol{\mu}_a), (Q_b, \boldsymbol{\mu}_b)$.*

*Proof.* $\Leftarrow$: First we show that if $g$ satisfies Ineq. 46 for given $f$, $g$ must be of the form in Eq. 48.

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:

$$g(\hat{Q}, \hat{\boldsymbol{\mu}}) = g(\mathbf{b}) + \sum_{i=1}^{n} g(\mathbf{x}_{n,i}) - g(\mathbf{x}_{n,i-1}). \tag{49}$$

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

$$\begin{aligned} & \sum_{i=1}^{n} \mathbf{x}_{n,i-1} (f(\mathbf{x}_{n,i}) - f(\mathbf{x}_{n,i-1})) && (50) \\ \leq \; & g(\hat{Q}, \hat{\boldsymbol{\mu}}) - C && (51) \\ \leq \; & \sum_{i=1}^{n} \mathbf{x}_{n,i} (f(\mathbf{x}_{n,i}) - f(\mathbf{x}_{n,i-1})) && (52) \end{aligned}$$

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 for the path integral of some function.[2] It is

$$\begin{aligned}
& \sum_{i=1}^{n} \mathbf{x}_{n,i} \left( f\left(\mathbf{x}_{n,i}\right) - f\left(\mathbf{x}_{n,i-1}\right) \right) && (53) \\
= & \sum_{i=1}^{n} \mathbf{x}_{n,i} f\left(\mathbf{x}_{n,i}\right) - \mathbf{x}_{n,i-1} f\left(\mathbf{x}_{n,i-1}\right) - \sum_{i=1}^{n} (\mathbf{x}_{n,i} - \mathbf{x}_{n,i-1}) f\left(\mathbf{x}_{n,i-1}\right) && (54) \\
= & (\hat{Q}, \hat{\boldsymbol{\mu}}) f(\hat{Q}, \hat{\boldsymbol{\mu}}) - \mathbf{b} f(\mathbf{b}) - \sum_{i=1}^{n} (\mathbf{x}_{n,i} - \mathbf{x}_{n,i-1}) f\left(\mathbf{x}_{n,i-1}\right). && (55)
\end{aligned}$$

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

$$\begin{aligned}
& \sum_{i=1}^{n} \mathbf{x}_{n,i-1} \left( f\left(\mathbf{x}_{n,i}\right) - f\left(\mathbf{x}_{n,i-1}\right) \right) && (56) \\
= & \sum_{i=1}^{n} \mathbf{x}_{n,i} f\left(\mathbf{x}_{n,i}\right) - \mathbf{x}_{n,i-1} f\left(\mathbf{x}_{n,i-1}\right) - \sum_{i=1}^{n} (\mathbf{x}_{n,i} - \mathbf{x}_{n,i-1}) f\left(\mathbf{x}_{n,i}\right) && (57) \\
= & (\hat{Q}, \hat{\boldsymbol{\mu}}) f(\hat{Q}, \hat{\boldsymbol{\mu}}) - \mathbf{b} f(\mathbf{b}) - \sum_{i=1}^{n} (\mathbf{x}_{n,i} - \mathbf{x}_{n,i-1}) f\left(\mathbf{x}_{n,i}\right). && (58)
\end{aligned}$$

By Lemma A.10,

$$\sum_{i=1}^{n} (\mathbf{x}_{n,i} - \mathbf{x}_{n,i-1}) f\left(\mathbf{x}_{n,i-1}\right) \underset{k\to\infty}{\rightarrow} \int_{0}^{(\hat{Q},\hat{\boldsymbol{\mu}})} f(\mathbf{x}) d\mathbf{x} \underset{n\to\infty}{\leftarrow} \sum_{i=1}^{n} (\mathbf{x}_{n,i} - \mathbf{x}_{n,i-1}) f\left(\mathbf{x}_{n,i}\right). \quad (59)$$

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.

$$g(\hat{Q}, \hat{\mu}) = C + (\hat{Q}, \hat{\mu}) f(\hat{Q}, \hat{\mu}) - \int_{0}^{\hat{Q},\hat{\mu}} f(x) dx, \quad (60)$$

as claimed.

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

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

[2]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.15.

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

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

□

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

### D.3 Proof of Lemma 5.3

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

*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 $s$. □

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

Consider the special case of differentiable DSRs. Since by Lemma A.5 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 5.4. 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).

**Lemma E.1.** *Let $s$ be a differentiable proper DSR. Then there are differentiable $f, g$ with $s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, y) = f_E(\hat{Q}, \hat{\boldsymbol{\mu}})_e + f_Y(\hat{Q}, \boldsymbol{\mu})_e y - g(\hat{Q}, \hat{\boldsymbol{\mu}})$.*

*Proof.* Lemma 4.1 shows that there are $f, g$ with $s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, y) = f_E(\hat{Q}, \hat{\boldsymbol{\mu}})_e + f_Y(\hat{Q}, \boldsymbol{\mu})_e y - g(\hat{Q}, \hat{\boldsymbol{\mu}})$. It is left to show that in particular we can choose $f, g$ that are differentiable. We do this by choosing a representation where $g = 0$. (As noted before, this is possible because otherwise we can define $\tilde{f}_E(\hat{Q}, \hat{\boldsymbol{\mu}})_e = f_E(\hat{Q}, \hat{\boldsymbol{\mu}})_e - g(\hat{Q}, \hat{\boldsymbol{\mu}})$ to obtain $s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, y) = \tilde{f}_E(\hat{Q}, \hat{\boldsymbol{\mu}})_e + f_Y(\hat{Q}, \boldsymbol{\mu})_e y)$.

Now notice that for all $\hat{Q}, \hat{\boldsymbol{\mu}}, e$, $s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, 1) - s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, 0) = f_Y(\hat{Q}, \hat{\boldsymbol{\mu}})_e$. Since the difference of two differentiable functions is differentiable, $f_Y$ is differentiable. Furthermore, $s(\hat{Q}, \hat{\boldsymbol{\mu}}, e, 0) = f_E(\hat{Q}, \hat{\boldsymbol{\mu}})_e$. Thus, $f_E$ must also be differentiable. □

**Proposition E.2.** *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.*

$$s(\hat{Q}, \hat{\boldsymbol{\mu}}, Q, \boldsymbol{\mu}) = f(\hat{Q}, \hat{\boldsymbol{\mu}})(Q, \boldsymbol{\mu}) - g(\hat{Q}, \hat{\boldsymbol{\mu}}) \tag{61}$$

*with*

$$(Q,\boldsymbol{\mu})\frac{d}{d\mu_e}f(Q,\boldsymbol{\mu}) \quad = \quad \frac{d}{d\mu_e}g(Q,\boldsymbol{\mu}) \tag{62}$$

$$(Q,\boldsymbol{\mu})\frac{d}{dQ(e)}f(Q,\boldsymbol{\mu}) \quad = \quad \frac{d}{dQ(e)}g(Q,\boldsymbol{\mu}) \tag{63}$$

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

We here use the derivative with respect to entries of $Q$ without much care. A technical issue here is that if we only change one entry of $Q$, we do not obtain a probability distribution. Thus, for defining the derivative w.r.t. $Q(e)$ we have to specify a way to counterbalance the change in $Q(e)$ to obtain a new probability distribution to feed into $f, g$. The definition of the derivative is ambiguous w.r.t. how we counterbalance. Since it doesn't matter for the proposition or proof how we specify the derivative, we do not have to resolve the ambiguity here.

*Proof.* $\Rightarrow$: By Lemma E.1, there are differentiable functions $f, g$ s.t. the expected scores are $s(\hat{Q},\hat{\boldsymbol{\mu}},Q,\boldsymbol{\mu}) = f(\hat{Q},\hat{\boldsymbol{\mu}})\boldsymbol{\mu} - g(\hat{Q},\hat{\boldsymbol{\mu}})$. 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.

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 62 and 63 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 62 and 63 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 62 and 63 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:

$$\frac{s'(p,1)}{s'(p,0)} = \frac{p-1}{p} \Leftrightarrow \frac{f'(p) - g'(p)}{g'(p)} = \frac{p-1}{p} \Leftrightarrow g'(p) = pf'(p), \tag{64}$$

which is exactly the relationship between $f', g'$ stated in our Proposition E.2. 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.2 be generalized to *non*-differentiable DSRs $s$, giving yet another style of characterization? At least on first sight, this seems difficult, because if $f$ can be discontinuous it is unclear what concept could replace the derivative of $f$.