Strictly Proper Contract Functions Can Be Arbitrage-Free

Eric Neyman, Tim Roughgarden

Columbia University
eric.neyman@columbia.edu, tim.roughgarden@gmail.com

Abstract

We consider mechanisms for truthfully eliciting probabilistic predictions from a group of experts. The standard approach — using a proper scoring rule to separately reward each expert — is not robust to collusion: experts may collude to misreport their beliefs in a way that guarantees them a larger total reward no matter the eventual outcome. (Chun and Shachter 2011) termed any such collusion "arbitrage" and asked whether there is any truthful elicitation mechanism that makes arbitrage impossible. We resolve this question positively, exhibiting a class of strictly proper arbitrage-free contract functions. These contract functions have two parts: one ensures that the total reward of a coalition of experts depends only on the average of their reports; the other ensures that changing this average report hurts the experts under at least one outcome.

1 Introduction

Suppose that some person or entity (the *principal*) wishes to elicit a probabilistic forecast from an expert: for example, a local news organization may want to ask a meteorologist for the probability that it will rain tomorrow. The principal could incentivize the expert with a payment scheme that depends on the expert's report (in our example, the meteorologist's stated probability) and the eventual outcome (whether or not it rains). Such a payment scheme is called a *scoring rule*, and a scoring rule is called *proper* if the expert's optimal strategy for maximizing expected payment is to report their belief. A scoring rule is called *strictly proper* if this is the expert's unique optimal strategy, no matter their belief. The space of all strictly proper scoring rules is very large, but the most well-known and frequently used one is Brier's quadratic scoring rule (Brier 1950), which gives an expert who reports p a reward of $1 - 2(1-p)^2$ if the event happens and $1 - 2p^2$ if it does not. This can be thought of as penalizing the expert based on their report's distance to the "right answer" (see Equation 1 in Section 3). The logarithmic scoring rule (Good 1952), which gives a reward of $\ln(p)$ if the event happens and $\ln(1 - p)$ if it does not, is also strictly proper and well known.

In many settings, the principal may wish to elicit forecasts from multiple experts, so as to get a better sense of

expert opinion and the extent to which there is a consensus. The principal could use the quadratic scoring rule (or any other strictly proper scoring rule) to elicit each expert's forecast. If experts are not allowed to collude, then this strategy is incentive-compatible; however, (French 1983) observed that experts can collude in a way that increases the sum total profit of all experts, no matter the final outcome. For example, if three experts believe that there is a 40%, 50%, and 90% chance of rain and are rewarded with the quadratic scoring rule, then their total reward is $0.28 + 0.5 + 0.98 = 1.76$ if it rains and $0.68 + 0.5 - 0.62 = 0.56$ if it does not; but if they collude to all report 60% then their reward is $0.68 + 0.68 + 0.68 = 2.04$ if it rains and $0.28 + 0.28 + 0.28 = 0.84$ if it does not. The experts can agree beforehand to a redistribution of their rewards in such a way that each expert is guaranteed to be better off than if they had not colluded.

(Chun and Shachter 2011) called this phenomenon — in which experts collude to misreport in a way where their total reward is larger no matter the outcome — *arbitrage*. They showed that every strictly proper scoring rule admits arbitrage — indeed, that there is an arbitrage opportunity for any group of experts so long as they do not all agree on the probability of the event. Specifically, a coalition of experts can risklessly make a profit by deviating to report an aggregate of their beliefs (in the case of the quadratic scoring rule, this aggregate is the arithmetic mean). (Neyman and Roughgarden 2021) extends this observation to probability distributions over more than two possible outcomes.

For many reasons, the expert may wish to make arbitrage impossible. First, the principal may wish to know whether the experts are in agreement: if they are not, for instance, the principal may want to elicit opinions from more experts. If the experts collude to report an aggregate value (as in our example), the principal does not find out whether they originally agreed. Second, even if the principal only seeks to act based on some aggregate of the experts' opinions, their method of aggregation may be different from the one that experts use to collude. For instance, the principal may have a private opinion on the trustworthiness of each expert and wishes to average the experts' opinions with corresponding weights. Collusion among the experts denies the principal this opportunity. Third, a principal may wish to track the accuracy of each individual expert (to figure out which experts

to trust more in the future, for instance), and collusion makes this impossible. Fourth, the space of collusion strategies that constitute arbitrage is large. In our example above, any report in $[0.546, 0.637]$ would guarantee a profit; and this does not even mention strategies in which experts report different probabilities. As such, the principal may not even be able to recover basic information about the experts' beliefs from their reports.

As a practical example, a government may wish to sponsor several competing teams to forecast the weather as accurately as possible, with the goal of hiring the best team. To provide good incentives, it pays each team using a proper scoring rule. A subset of the teams realize that they would benefit from colluding: for example, if the quadratic scoring rule is used, the experts may collude to all report the average of their true predictions. This would interfere with the government's goal of finding the best team, so using an arbitrage-proof mechanism would be desirable.

As another example, the U.S. Federal Reserve may wish to elicit economic growth forecasts from several experts. The Federal Reserve wants to know not just the experts' consensus, but also how much the experts disagree: perhaps, if the experts disagree substantially, they would want to pursue the disagreement further and try to resolve it. If the experts collude to report the same forecast, then the federal reserve will not find out whether the experts disagree. Thus, it may wish to instead use a arbitrage-proof mechanism to reward the experts.

As we have discussed, preventing arbitrage is impossible if the principal chooses a strictly proper scoring rule and uses it to reward all of the experts. However, the principal has more freedom than this: they may choose to make each expert's reward depend not only on that expert's report but also *other* experts' reports. (Chun and Shachter 2011, §5) ask whether there is any mechanism for rewarding experts that makes arbitrage impossible, concluding that this "seems unlikely". (Freeman et al. 2020) explore this question further, proposing a mechanism that prevents arbitrage if the experts' reports are guaranteed to be in the range $[\epsilon, 1-\epsilon]$ for some positive ϵ (though their mechanism may require very large payments if ϵ is small). However, they leave open Chun and Shachter's question of whether an incentive-compatible, arbitrage-free reward mechanism exists.

We resolve this question in the affirmative by exhibiting a class of incentive-compatible mechanisms in which arbitrage from collusion is impossible. Our mechanism takes inspiration from Brier's quadratic scoring rule, but modifies it to take into account the aggregate performance of the remaining experts.

2 Related Work

(Freeman et al. 2020) explores the question of whether strictly proper arbitrage-free mechanisms exist by proving positive results under different relaxations of these constraints. Their main result is a strictly proper arbitrage-free mechanism under the restriction that the range of allowed reports is restricted to $[\epsilon, 1-\epsilon]$. However, their mechanism necessitates payments that are exponentially large in $1/\epsilon$. Al-

ternatively, these payments can be scaled down, but at the expense of giving essentially zero reward to each expert on the vast majority of the interval of possible reports, thus providing little incentive for truthful reporting. They also exhibit a positive result if the properness criterion is somewhat relaxed to allow for some contract functions that are proper but not strictly proper.

(Chen et al. 2014) explore the different but related topic of *arbitrage-free wagering mechanisms*. In a wagering mechanism, each expert wagers a certain amount of money along with their report, and the pool of wagers is redistributed among the experts depending on each expert's report and wager and the eventual outcome. In this setting, they define arbitrage as any opportunity for an individual to risklessly make a profit. That is, an arbitrage opportunity is one in which an expert may unilaterally deviate by submitting a report that guarantees a profit no matter the final outcome. This differs from Chun and Shachter's definition of arbitrage, which is concerned with riskless profit opportunities stemming from collusion between experts.

The most well-known wagering mechanism is the *weighted score wagering mechanism*, which rewards each expert based on their performance compared to other experts according to a strictly proper scoring rule. An expert may risklessly profit from a weighted score wagering mechanism by reporting an aggregate of other experts' reports. This is the same aggregate as the one that a coalition of experts who are rewarded with a strictly proper scoring rule may report in order to risklessly make a profit in our setting. Chen et al. define *no-arbitrage wagering mechanisms*, which modify the reallocation rule of weighted score wagering mechanisms to reward each expert based on their performance relative to the performance of the aggregate of all other experts' reports (Chen et al. 2014, §4.1). No-arbitrage wagering mechanisms can be re-interpreted in our setting as contract functions that prevent the *entire group of experts* from colluding. However, as we discuss in Section 3, this is easy to accomplish; we are instead faced with the challenge of preventing collusion between any coalition of experts of any size. Thus, our mechanism and theirs share some of the same spirit, but are different mechanisms that solve different problems.

Finally, we note some superficial resemblances (e.g., the use of proper scoring rules as a useful primitive) between the present line of work and peer prediction (Miller, Resnick, and Zeckhauser 2005), an area that studies how to solicit honest opinions of subjective events. Contract functions concern verifiable events and thus are much more closely related to (generalizations of) scoring rules than to peer prediction.

3 Preliminaries

We consider an event with n possible outcomes. The space of possible probability distributions over these outcomes is Δ_n, the standard simplex in $\mathbb{R}^n$. We denote the vertices of this simplex by δ_j for $j \in [n]$ (δ_j is the vector whose j-th coordinate is 1 and all of whose other coordinates are 0).

Scoring rules A scoring rule is any function that takes as input a probability distribution over the n possible outcomes, and the eventual outcome, and outputs a reward. Formally, a scoring rule is any function $s : \Delta_n \times [n] \to \mathbb{R}$; if an expert reports probability distribution $\mathbf{p} \in \Delta_n$ and the outcome is j, then the expert receives reward $s(\mathbf{p}; j)$.

A scoring rule is *proper* if an expert with belief $\mathbf{b}$ maximizes their expected reward by reporting $\mathbf{b}$. Formally, s is proper if for all $\mathbf{b}$, $\sum_j b_j s(\mathbf{x}; j)$ is maximized at $\mathbf{x} = \mathbf{b}$. We say that s is *strictly proper* if for all $\mathbf{b}$, $\mathbf{x} = \mathbf{b}$ is the unique maximizer, i.e. that an expert does strictly worse by misreporting their belief.

Brier's quadratic scoring rule — mentioned in the introduction — is the scoring rule

$$s_{\text{quad}}(\mathbf{p}; j) := 1 - (1 - p_j)^2 - \sum_{\ell \neq j} p_\ell^2.$$

It can be rewritten as

$$s_{\text{quad}}(\mathbf{p}; j) = 1 - \|\mathbf{p} - \delta_j\|_2^2. \qquad (1)$$

Thus, the quadratic scoring rule can be thought of as penalizing the expert by the squared distance between their report and the "omniscient" answer δ_j. The quadratic scoring rule is strictly proper. In the case of $n = 2$ outcomes, the quadratic scoring rule can be written[1] as

$$s_{\text{quad}}(\mathbf{p}; j) = 1 - 2(1 - p_j)^2.$$

Contract functions Contract functions, defined by (Chun and Shachter 2011), generalize scoring rules to multiple experts. We say that there are m experts; for $i \in [m]$, expert i reports a probability distribution $\mathbf{p}_i \in \Delta_n$. We denote the j-th coordinate of $\mathbf{p}_i$ as $p_{i,j}$.

A *contract function* is any function that takes as input the m experts' reports and the outcome, and outputs the reward of each expert. Formally, a contract function is any function $\Pi : (\Delta_n)^m \times [n] \to \mathbb{R}^m$; if the experts report distributions $\mathbf{p}_1, \ldots, \mathbf{p}_m$ and the outcome is j, then the vector of expert rewards is $\Pi(\mathbf{p}_1, \ldots, \mathbf{p}_m; j)$. We let $\Pi_i(\cdot)$ denote the i-th coordinate of $\Pi(\cdot)$, i.e. expert i's reward. We will generally use $\mathbf{P}$ to denote the m-tuple of reports $(\mathbf{p}_1, \ldots, \mathbf{p}_m)$.

A contract function is *proper* if for each $i \in [m]$, expert i maximizes their expected reward by reporting their belief $\mathbf{b}_i$, no matter the reports $\mathbf{p}_{-i}$ of the other experts. Formally, Π is proper if for all $i \in [m]$, for all $\mathbf{b}_i$ and all $\mathbf{p}_{-i}$, $\sum_j b_{i,j} \Pi_i(\mathbf{x}, \mathbf{p}_{-i}; j)$ is maximized at $\mathbf{x} = \mathbf{b}_i$. We say that Π is *strictly proper* if $\mathbf{x} = \mathbf{b}_i$ is the unique maximizer, i.e. that an expert does strictly worse by misreporting their belief.

Our goal is to exhibit a strictly proper contract function that does not permit arbitrage from collusion. We use the definition of arbitrage given by (Freeman et al. 2020), which was adapted from (Chun and Shachter 2011).

A contract function Π *admits arbitrage* if there is a coalition (i.e. subset) $C \subseteq [m]$ of experts and m-tuples of expert reports $\mathbf{P}$ and $\mathbf{Q}$, with $\mathbf{p}_i = \mathbf{q}_i$ for all $i \notin C$, such that

$$\sum_{i \in C} \Pi_i(\mathbf{Q}; j) \geq \sum_{i \in C} \Pi_i(\mathbf{P}; j)$$

for all $j \in [n]$, and the inequality is strict for some j. We say that Π is *arbitrage-free* if it does not admit arbitrage. Intuitively, Π admits arbitrage if it is possible for a coalition of experts to collude to misreport their values in such a way that the total reward of the experts in the coalition ends up larger, no matter the outcome. (Above, the misreport is $\mathbf{Q}$; the constraint that $\mathbf{p}_i = \mathbf{q}_i$ for $i \notin C$ means that only experts in C change their reports.) If this is possible, then the experts in C can commit beforehand to a redistribution of the extra reward in a way that makes every expert in the coalition better off no matter the eventual outcome j.

Remark 3.1. Positive affine transformations preserve both strict properness and arbitrage-freeness. That is, if Π is strictly proper then so is $a\Pi + b$ for any $a > 0$ and b, and this is likewise true for arbitrage-freeness.

The question posed by (Chun and Shachter 2011) and explored by (Freeman et al. 2020), which we answer affirmatively in this work, is: **Does there exist a strictly proper arbitrage-free contract function?**

In the case of $m = 2$ experts, there is a fairly straightforward solution:

$$\Pi(\mathbf{p}_1, \mathbf{p}_2; j) = (s_{\text{quad}}(\mathbf{p}_1; j) - s_{\text{quad}}(\mathbf{p}_2; j), s_{\text{quad}}(\mathbf{p}_2; j) - s_{\text{quad}}(\mathbf{p}_1; j)).$$

$$(2)$$

This contract function is strictly proper because expert 1's reward is the (strictly proper) quadratic score of their report plus a term that does not depend on their report, and likewise for expert 2. It is arbitrage-free because the total reward of the two experts is 0 no matter what. Indeed, this contract function is arbitrage-free with any strictly proper scoring rule in place of the quadratic scoring rule.

This idea does not extend to $m > 2$ experts, because an arbitrage-free contract function must not admit arbitrage by a coalition of experts of any size. While it is easy to construct a contract function that does not admit arbitrage by a coalition of size m (by making the total reward always equal to 0), this does not automatically make the contract function free of arbitrage opportunities for coalitions of sizes between 2 and $m-1$. In the next section we address this challenge and exhibit a strictly proper contract function that is arbitrage-free for $m > 2$ experts.

4 A Class of Strictly Proper Arbitrage-Free Contract Functions

Suppose that — as before — there are $m \geq 2$ experts who are forecasting an event with $n \geq 2$ outcomes. Given experts with reports $\mathbf{P} = (\mathbf{p}_1, \ldots, \mathbf{p}_m)$ and a nonempty subset $S \subseteq [m]$ of the experts, we will let $\overline{\mathbf{p}}_S := \frac{1}{|S|} \sum_{i \in S} \mathbf{p}_i$ be the average of the experts' reports. We will use $\overline{\mathbf{p}}_{-i}$ to denote $\overline{\mathbf{P}}_{[m] \setminus \{i\}}$.

We now state our main theorem, which exhibits a class of strictly proper, arbitrage-free contract functions.

Theorem 4.1. *Let α be a real number such that $\alpha < 0$ or $\alpha \geq 2(m-1)^2 n$. Let Π be the contract function defined by*

$$\Pi_i(\mathbf{P};j) = s_{quad}(\mathbf{p}_i;j) - (m-1)^2 s_{quad}(\overline{\mathbf{p}}_{-i};j) + \alpha \overline{\mathbf{p}}_{-i,j}$$

for each i, j. Then Π is strictly proper and arbitrage-free.

Note that in the case of $m = 2$, setting $\alpha = 0$ yields our aforementioned solution for two experts in Equation 2. Unfortunately, setting $\alpha = 0$ for $m > 2$ experts causes arbitrage-freeness to fail in certain edge cases.

One can think of the contract function in Theorem 4.1 as having two parts. The first part, $s_{\text{quad}}(\mathbf{p}_i;j) - (m-1)^2 s_{\text{quad}}(\overline{\mathbf{p}}_{-i};j)$, ensures that any coalition's total reward depends only on the average of the coalition's reports. In effect this significantly limits the degrees of freedom that a coalition has when colluding. The second part, $\alpha \overline{\mathbf{p}}_{-i,j}$, ensures that any deviation in this average report causes a decrease in total reward under at least one outcome.

In this section we focus on proving Theorem 4.1 for $n = 2$ outcomes, as this allows us to simplify notation without sacrificing the core ideas. We defer the proof for general values of n to the appendix.

Proof of Theorem 4.1 for $n = 2$. First, note that Π is strictly proper, because expert i's reward is their quadratic score plus a term that does not depend on their report. It remains to show that Π is arbitrage-free.

Let $C \subseteq [m]$ be a coalition of experts. Strict properness entails that no expert can unilaterally find an arbitrage opportunity, so we may assume that $|C| \geq 2$.

For an outcome j and a subset $S \subseteq [m]$, let $p_{S,j} := \sum_{i \in S} p_{i,j}$. The following fact follows from algebraic manipulations, which we defer to the appendix.

Lemma 4.2. *The expression for $\Pi_i(\mathbf{P};j)$ is equal to*

$$2(p_{[m],j} - d - 1)(p_{[m],j} - 2p_{i,j} - d + 1) + f(m,\alpha), \quad (3)$$

for some function f, where $d = m - 1 - \frac{\alpha}{4(m-1)}$.

Equation 3 makes it evident that rewards add nicely across experts in a coalition C, as the first term of the product is the same for all experts in C. We will use the notation $\Pi_C(\mathbf{P};j)$ to denote $\sum_{i \in C} \Pi_i(\mathbf{P};j)$. The key idea is that, as we are about to show, if the reports of experts not in C are held fixed, $\Pi_C(\mathbf{P};j)$ depends *only* on $p_{C,j}$. Thus, the experts in C have only one degree of freedom available for colluding: the sum of their reports.

We write $\overline{C}$ to mean $[m] \setminus C$. We have

$$\Pi_C(\mathbf{P};j) = 2\sum_{i \in C}(p_{[m],j} - d - 1)(p_{[m],j} - 2p_{i,j} - d + 1)$$
$$+ |C| f(m,\alpha)$$
$$= 2(p_{C,j} + p_{\overline{C},j} - d - 1)$$
$$\cdot ((|C| - 2)p_{C,j} + |C|(p_{\overline{C},j} - d + 1)) + |C| f(m,\alpha)$$
$$= 2((|C| - 2)p_{C,j}^2 + 2((|C| - 1)(p_{\overline{C},j} - d) + 1)p_{C,j})$$
$$+ g(m,\alpha,|C|,p_{\overline{C},j}), \quad (4)$$

for some function g. Now, recall the constraints on α in Theorem 4.1, and note that $\alpha < 0 \Leftrightarrow d > m - 1$ and $\alpha \geq 4(m-1)^2 \Leftrightarrow d \leq 0$. With this in mind, we now prove the following claim, which is sufficient to complete our proof.

Claim 4.3. *If $d \leq 0$, then for each j and for all possible reports of experts not in C, $\Pi_C(\mathbf{P};j)$ is a strictly increasing function of $p_{C,j}$. If $d > m - 1$, it is a strictly decreasing function of $p_{C,j}$.*

By virtue of deriving Equation 4, we have already proven the most difficult part of Claim 4.3, which is that $\Pi_C(\mathbf{P};j)$ is a function of (i.e. determined by) $p_{C,j}$. Why is this function's monotonicity sufficient to complete our proof of Theorem 4.1? Since $p_{C,1} + p_{C,2} = |C|$, it follows from Claim 4.3 that for $d \leq 0$ and $d > m - 1$, colluding in a way that increases the total reward in the case of one outcome necessarily decreases it in the case of the other outcome.

Proof of Claim 4.3. We first consider the case of $|C| = 2$. In this case we have

$$\Pi_C(\mathbf{P};j) = 4(p_{\overline{C},j} - d + 1)p_{C,j} + g(m,\alpha,|C|,p_{\overline{C},j}).$$

Now, $0 \leq p_{\overline{C},j} \leq m - 2$, which means that $1 - d \leq p_{\overline{C},j} - d + 1 \leq m - 1 - d$. If $d \leq 0$, this quantity is guaranteed to be strictly positive, so $\Pi_C(\mathbf{P};j)$ is a strictly increasing function of $p_{C,j}$; if $d > m - 1$, it is guaranteed to be negative, so $\Pi_C(\mathbf{P};j)$ is a strictly decreasing function of $p_{C,j}$.

Now assume that $|C| > 2$. In this case, it follows from Equation 4 that $\Pi_C(\mathbf{P};j)$ is a parabola with a minimum at

$$\frac{(|C| - 1)(d - p_{\overline{C},j}) - 1}{|C| - 2}.$$

We wish to show that if $d \leq 0$ then this quantity is at most 0, and that if $d > m - 1$ then it is at least $|C|$ (since the range of possible values of $p_{C,j}$ is $[0,|C|]$). If $d \leq 0$ then, since $p_{\overline{C},j} \geq 0$, we have

$$\frac{(|C| - 1)(d - p_{\overline{C},j}) - 1}{|C| - 2} \leq \frac{-1}{|C| - 2} \leq \frac{-1}{m - 2} \leq 0.$$

If $d > m - 1$ then, since $p_{\overline{C},j} \leq m - |C|$, we have

$$\frac{(|C| - 1)(d - p_{\overline{C},j}) - 1}{|C| - 2} \geq \frac{(|C| - 1)^2 - 1}{|C| - 2} = |C|.$$

$\square$

Having proved the claim, we have completed the proof of Theorem 4.1 for $n = 2$. $\square$

We note that setting $\alpha = 0$ results in a contract function that is arbitrage-free except in one edge case: in the event that all but two experts assign a probability of zero to some outcome j, the remaining experts can collude to adjust their probabilities — in particular, lowering the total probability they assign to outcome j — in a way that increases

their total reward under outcome j and leaves the remaining rewards unchanged. If we are willing to put this exception aside (e.g. if we only allow reports strictly between 0 and 1), then we may regard the resulting contract function $\Pi_i(\mathbf{P};j) = s_{\text{quad}}(\mathbf{p}_i;j) - (m-1)^2 s_{\text{quad}}(\overline{\mathbf{P}}_{-i};j)$ as arbitrage-free. This contract function has a natural interpretation: it rewards an expert for the accuracy of their forecast but penalizes the expert if others are accurate in aggregate. This rule is reminiscent of the no-arbitrage wagering mechanism for the quadratic scoring rule given in (Chen et al. 2014), except that the penalty is multiplied by a factor of $(m-1)^2$.

5 Future Directions in Arbitrage-Freeness

Having exhibited a strictly proper arbitrage-free contract function, it is natural to ask about other (possibly stronger) notions of arbitrage-freeness. Another natural notion is to say that Π admits arbitrage if a coalition C of experts can collude in a way that, in the opinion of every expert in C, increases the expected total reward of the experts in C. Formally:

Definition 5.1. *A contract function Π admits expected arbitrage if there is a coalition $C \subseteq [m]$ of experts and vectors of reports $\mathbf{P} = (\mathbf{p}_1, \ldots, \mathbf{p}_m)$, $\mathbf{Q} = (\mathbf{q}_1, \ldots, \mathbf{q}_m)$, with $p_i = q_i$ if $i \notin C$, such that for all $i \in C$ we have*

$$\sum_{j \in [n]} p_{i,j} \sum_{k \in C} \Pi_k(\mathbf{P};j) \leq \sum_{j \in [n]} p_{i,j} \sum_{k \in C} \Pi_k(\mathbf{Q};j),$$

and the inequality is strict for some i. We say that Π is free of expected arbitrage if it does not admit expected arbitrage.

Up to edge scenarios,[2] if a contract function admits arbitrage then it also admits expected arbitrage. On the other hand, in the case of $m > 2$ experts, the scoring rules described by Theorem 4.1 (which do not admit arbitrage) do admit expected arbitrage. As an example, consider two outcomes and m experts with beliefs $(\frac{1}{2}, \frac{1}{2})$. If all experts report their beliefs, then each expert's reward is $\frac{\alpha}{2} + \frac{1}{2}(1 - (m-1)^2)$ no matter the outcome. If all experts instead report $(1,0)$ then each expert expects a reward of $\frac{\alpha}{2}$, which is larger. This raises the following question.

Question 5.2. *Is there a strictly proper scoring rule that does not admit expected arbitrage?*

We hope that our work will spur research on stronger notions of arbitrage-freeness, included but not limited to Definition 5.1.

A Details Omitted from Section 4

Proof of Lemma 4.2. Let $d = m - 1 - \frac{\alpha}{4(m-1)}$. We use f_0 and f to denote functions whose particular values do not concern us. We have

$$\Pi_i(\mathbf{P};j) = s_{\text{quad}}(\mathbf{p}_i;j) - (m-1)^2 s_{\text{quad}}(\overline{\mathbf{P}}_{-i};j) + \alpha \overline{\mathbf{p}}_{-i,j}$$

---

[2]It is possible for a coalition of experts to collude in a way that increases their total reward under an outcome to which they all assign probability 0. If their reward in the case of all other outcome is unchanged, such a deviation would constitute arbitrage but not expected arbitrage.

$$= 1 - 2(1 - p_{i,j})^2 - (m-1)^2 \left(1 - 2\left(1 - \frac{p_{-i,j}}{m-1}\right)^2\right)$$
$$+ \frac{\alpha}{m-1} p_{-i,j}$$
$$= -2(1 - p_{i,j})^2 + 2(m - 1 - p_{-i,j})^2 + \frac{\alpha}{m-1} p_{-i,j} + f_0(m, \alpha)$$
$$= -2(1 - p_{i,j})^2 + 2(d - p_{-i,j})^2 + f(m, \alpha)$$
$$= 2(d - p_{-i,j} + (1 - p_{i,j}))(d - p_{-i,j} - (1 - p_{i,j})) + f(m, \alpha)$$
$$= 2(p_{[m],j} - d - 1)(p_{[m],j} - 2p_{i,j} - d + 1) + f(m, \alpha),$$

as desired. $\qquad\square$

Proof of Theorem 4.1 for general n. First, note that Π is strictly proper, because expert i's reward is their quadratic score plus a term that does not depend on their report. It remains to show that Π is arbitrage-free.

Let $C \subseteq [m]$ be a coalition of experts. Strict properness entails that no expert can unilaterally find an arbitrage opportunity, so we may assume that $|C| \geq 2$.

For an outcome j and a subset $S \subseteq [m]$, let $p_{S,j} := \sum_{i \in S} p_{i,j}$. Let $d = m - 1 - \frac{\alpha}{2(m-1)}$. We rewrite $\Pi_i(\mathbf{P};j)$ as follows, using f_0 and f to denote functions whose particular values do not concern us.

$$\Pi_i(\mathbf{P};j) = s_{\text{quad}}(\mathbf{p}_i;j) - (m-1)^2 s_{\text{quad}}(\overline{\mathbf{P}}_{-i};j) + \alpha \overline{\mathbf{p}}_{-i,j}$$
$$= (m-1)^2 \left(1 - \frac{1}{m-1} p_{-i,j}\right)^2 - (1 - p_{i,j})^2$$
$$+ \sum_{\ell \neq j} \left((m-1)^2 \left(\frac{1}{m-1} p_{-i,\ell}\right)^2 - p_{i,\ell}^2\right)$$
$$+ \frac{\alpha}{m-1} p_{-i,j} + f_0(m, n, \alpha)$$
$$= (d - p_{-i,j})^2 - (1 - p_{i,j})^2$$
$$+ \sum_{\ell \neq j} (p_{-i,\ell}^2 - p_{i,\ell}^2) + f(m, n, \alpha)$$
$$= (d - p_{-i,j} + (1 - p_{i,j}))(d - p_{-i,j} - (1 - p_{i,j}))$$
$$+ \sum_{\ell \neq j} (p_{-i,\ell} + p_{i,\ell})(p_{-i,\ell} - p_{i,\ell}) + f(m, n, \alpha)$$
$$= (p_{[m],j} - d - 1)(p_{[m],j} - 2p_{i,j} - d + 1)$$
$$+ \sum_{\ell \neq j} p_{[m],\ell}(p_{[m],\ell} - 2p_{i,\ell}) + f(m, n, \alpha).$$

We will use the notation $\Pi_C(\mathbf{P};j)$ to denote $\sum_{i \in C} \Pi_i(\mathbf{P};j)$. We also write $\overline{C}$ to mean $[m] \setminus C$ and $\mathbf{P}_{\overline{C}}$ to mean the collection of reports $\mathbf{p}_i$ for $i \in \overline{C}$. We have

$$\Pi_C(\mathbf{P};j) = \sum_{i \in C} \Bigg((p_{[m],j} - d - 1)(p_{[m],j} - 2p_{i,j} - d + 1)$$
$$+ \sum_{\ell \neq j} p_{[m],\ell}(p_{[m],\ell} - 2p_{i,\ell}) \Bigg) + |C| f(m, n, \alpha)$$
$$= (p_{C,j} + p_{\overline{C},j} - d - 1)((|C| - 2)p_{C,j} + |C|(p_{\overline{C},j} - d + 1))$$
$$+ \sum_{\ell \neq j} (p_{C,\ell} + p_{\overline{C},\ell})(|C| p_{\overline{C},\ell} + (|C| - 2)p_{C,\ell}) + |C| f(m, n, \alpha)$$
$$= (|C| - 2)p_{C,j}^2 + ((2|C| - 2)(p_{\overline{C},j} - d) + 2)p_{C,j}$$

$$+ \sum_{\ell \neq j} \left((|C| - 2)p_{C,\ell}^2 + (2|C| - 2)p_{\overline{C},\ell}p_{C,\ell} \right) + g(m, n, \alpha, |C|, \mathbf{P}_{\overline{C}})$$

$$= (2 - (2|C| - 2)d)p_{C,j}$$
$$+ \sum_{\ell} \left((|C| - 2)p_{C,\ell}^2 + (2|C| - 2)p_{\overline{C},\ell}p_{C,\ell} \right) + g(m, n, \alpha, |C|, \mathbf{P}_{\overline{C}})$$

for some function g. Consider a different vector $\mathbf{Q}$ that agrees with $\mathbf{P}$ on $\overline{C}$.

Case 1: $\alpha < 0$. In this case, $2 - (2|C| - 2)d < 2 - (2|C| - 2)(m-1)$. Let $\tilde{j} = \arg\max_{\ell} q_{C,\ell} - p_{C,\ell}$, and let $\epsilon = q_{C,\tilde{j}} - p_{C,\tilde{j}}$. We note that

$$\sum_{\ell} q_{C,\ell}^2 - p_{C,\ell}^2 = \sum_{\ell}(q_{C,\ell} - p_{C,\ell})(q_{C,\ell} + p_{C,\ell})$$
$$\leq \epsilon \sum_{\ell}(q_{C,\ell} + p_{C,\ell}) = 2\epsilon|C|.$$

Thus, we have

$$\Pi_C(\mathbf{Q}; \tilde{j}) - \Pi_C(\mathbf{P}; \tilde{j}) = (2 - (2|C| - 2)d)\epsilon$$
$$+ \sum_{\ell}(|C| - 2)(q_{C,\ell}^2 - p_{C,\ell}^2) + (2|C| - 2)(q_{C,\ell} - p_{C,\ell})p_{\overline{C},\ell}$$
$$\leq (2 - (2|C| - 2)d)\epsilon + (|C| - 2) \cdot 2\epsilon|C|$$
$$+ (2|C| - 2)(m - |C|)\epsilon$$
$$\leq (2 - (2|C| - 2)(m-1))\epsilon + (|C| - 2) \cdot 2\epsilon|C|$$
$$+ (2|C| - 2)(m - |C|)\epsilon$$
$$= 2\epsilon(1 + (|C| - 2)|C| + (|C| - 1)(1 - |C|)) = 0,$$

with equality in the second step only when $\epsilon = 0$, i.e. $q_{C,\ell} = p_{C,\ell}$ for all ℓ. Thus, either the total reward of the experts in C is the same under $\mathbf{Q}$ as under $\mathbf{P}$ for every outcome, or it is strictly smaller under $\mathbf{Q}$ in the case of outcome $\tilde{j}$.

Case 2: $\alpha \geq 2(m-1)^2 n$. In this case, $2 - (2|C| - 2)d \geq 2 + (2|C| - 2)(m-1)(n-1)$. $\tilde{j} = \arg\max_{\ell} p_{C,\ell} - q_{C,\ell}$, and let $\epsilon = p_{C,\tilde{j}} - q_{C,\tilde{j}}$. Since $\sum_{\ell}(q_{C,\ell} - p_{C,\ell}) = 0$, it follows that $q_{C,\ell} - p_{C,\ell} \leq (n-1)\epsilon$ for all ℓ. We note that

$$\sum_{\ell} q_{C,\ell}^2 - p_{C,\ell}^2 \leq \sum_{\ell}(q_{C,\ell} + p_{C,\ell})\max(q_{C,\ell} - p_{C,\ell}, 0)$$
$$\leq 2|C|\sum_{\ell}\max(q_{C,\ell} - p_{C,\ell}, 0) \leq 2|C|(n-1)\epsilon.$$

We also have that

$$\sum_{\ell}(q_{C,\ell} - p_{C,\ell})p_{\overline{C},\ell} \leq (m - |C|)\sum_{\ell}\max(q_{C,\ell} - p_{C,\ell}, 0)$$
$$\leq (m - |C|)(n-1)\epsilon.$$

Therefore,

$$\Pi_C(\mathbf{Q}; \tilde{j}) - \Pi_C(\mathbf{P}; \tilde{j})$$
$$= -(2 - (2|C| - 2)d)\epsilon + \sum_{\ell}(|C| - 2)(q_{C,\ell}^2 - p_{C,\ell}^2)$$
$$+ (2|C| - 2)(q_{C,\ell} - p_{C,\ell})p_{\overline{C},\ell}$$
$$\leq -(2 - (2|C| - 2)d)\epsilon + (|C| - 2) \cdot 2|C|(n-1)\epsilon$$
$$+ (2|C| - 2) \cdot (m - |C|)(n-1)\epsilon$$
$$= 2\epsilon(-1 - (|C| - 1)(m-1)(n-1)$$
$$+ (n-1)(m(|C| - 1) - |C|))$$

$$= 2\epsilon(-1 - (n-1)) = -2\epsilon n \leq 0,$$

with equality in the last step only when $\epsilon = 0$, i.e. $q_{C,\ell} = p_{C,\ell}$ for all ℓ. As in the previous case, this means that either the total reward of the experts in C is the same under $\mathbf{Q}$ as under $\mathbf{P}$ for every outcome, or it is strictly smaller under $\mathbf{Q}$ in the case of outcome $\tilde{j}$. This completes the proof. $\square$

References

Brier, G. W. 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78: 1–3.

Chen, Y.; Devanur, N. R.; Pennock, D. M.; and Vaughan, J. W. 2014. Removing arbitrage from wagering mechanisms. In Babaioff, M.; Conitzer, V.; and Easley, D. A., eds., *ACM Conference on Economics and Computation, EC '14, Stanford, CA, USA, June 8-12, 2014*, 377–394. ACM.

Chun, S.; and Shachter, R. D. 2011. Strictly Proper Mechanisms with Cooperating Players. In Cozman, F. G.; and Pfeffer, A., eds., *UAI 2011, Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence, Barcelona, Spain, July 14-17, 2011*, 125–134. AUAI Press.

Freeman, R.; Pennock, D. M.; Peters, D.; and Waggoner, B. 2020. Preventing Arbitrage from Collusion When Eliciting Probabilities. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, 1958–1965. AAAI Press.

French, S. 1983. *Group Consensus Probability Distributions: A Critical Survey*. University of Manchester. Department of Decision Theory.

Good, I. J. 1952. Rational Decisions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 14(1): 107–114.

Miller, N.; Resnick, P.; and Zeckhauser, R. 2005. Eliciting Informative Feedback: The Peer-Prediction Method. *Manag. Sci.*, 51(9): 1359–1373.

Neyman, E.; and Roughgarden, T. 2021. From Proper Scoring Rules to Max-Min Optimal Forecast Aggregation. In Biró, P.; Chawla, S.; and Echenique, F., eds., *ACM Conference on Economics and Computation, EC '21, Budapest, Hungary, July 18-23, 2021*, 734. ACM.