

# Preventing Arbitrage from Collusion When Eliciting Probabilities

Rupert Freeman,<sup>1</sup> David M. Pennock,<sup>1</sup> Dominik Peters,<sup>2</sup> Bo Waggoner<sup>3</sup>

<sup>1</sup>Microsoft Research, <sup>2</sup>Carnegie Mellon University, <sup>3</sup>CU Boulder

{rupert.freeman, dpennock}@microsoft.com, dominikp@cs.cmu.edu, bwag@colorado.edu

## Abstract

We consider the design of mechanisms to elicit probabilistic forecasts when agents are strategic and may collude with one another. Chun and Shachter (2011) have shown that when agents may form coalitions, many known mechanisms for elicitation permit arbitrage, allowing the coalition members to guarantee themselves higher payments by misreporting their beliefs. We consider two approaches to protect against colluding agents. First, we present a novel strictly proper mechanism that does not admit arbitrage provided that the reports of the agents are bounded away from 0 and 1, a common assumption in many settings. Second, we discover strictly arbitrage-free mechanisms that satisfy an intermediate guarantee between weak and strict properness.

## 1 Introduction

Suppose that a decision maker, the *principal*, asks an agent to estimate the probability that it will rain tomorrow. The principal would like the agent to respond with their true belief. The principal can incentivize truthfulness by asking the agent for the prediction today, then paying the agent an amount of money tomorrow that depends on whether it actually rained and how accurate the prediction was. Payment schemes in this setting are called *scoring rules*. For example, using Brier’s (1950) famous quadratic scoring rule, the principal would pay the agent  $\$1 - (1 - \hat{p})^2$  if it rains, and  $\$1 - \hat{p}^2$  if it does not, where  $\hat{p}$  is the probability that the agent reported. For example, if  $\hat{p} = 0.6$ , then the agent earns  $\$0.84$  if it rains and  $\$0.64$  otherwise. Calculus shows that if the agent wishes to maximize their expected payment, it is uniquely optimal for the agent to report their truthful subjective belief. Scoring rules with this property are called *strictly proper*.

The principal may wish to obtain an estimate from several agents. If each agent is paid their quadratic score, every agent is incentivized to report truthfully, just as in the single-agent case. However, French (1985) observed that the quadratic score may not be incentive compatible if agents are able to collude. Assume that the agents know each other, can communicate their beliefs, and can transfer money among themselves. Suppose three agents were to truthfully report  $\hat{\mathbf{p}} = (0.2, 0.4, 0.6)$ . Then the quadratic score pays them a total of  $\$0.36 + \$0.64 + \$0.84 = \$1.84$

if it rains, and  $\$0.96 + \$0.84 + \$0.64 = \$2.44$  otherwise. If the group instead decided to calculate their mean belief (0.4) and report  $\hat{\mathbf{p}} = (0.4, 0.4, 0.4)$ , then quadratic score payments are  $\$0.64 + \$0.64 + \$0.64 = \$1.92$  if it rains, and  $\$0.84 + \$0.84 + \$0.84 = \$2.52$  otherwise. In both cases, the group obtains strictly more money than before. They could then distribute their gains in some way so that each member is strictly better off, no matter the outcome. This example is general: groups with at least some disagreement are always strictly better off if they report their mean belief. French (1985) shows that this is true not only for Brier’s quadratic rule, but for every concave scoring rule. Chun and Shachter (2011) termed this phenomenon *arbitrage*, since it allows groups of colluding agents to *risklessly* increase their payoff. They showed that *all* strictly proper scoring rules, whether concave or not, permit arbitrage, and that more complicated multi-agent rules are also vulnerable, including market scoring rules (Hanson 2003) and competitive scoring rules (Kilgour and Gerchak 2004; Lambert et al. 2008).

We may hope that collusion of this type is uncommon due to coordination and communication difficulties. However, it is often the case that forecasters are assigned to small groups for the purposes of deliberation before reporting their probabilities.<sup>1</sup> While deliberation of this type can improve forecast accuracy by encouraging information sharing, it also provides an opportunity for collusion. Further, the profit opportunity might encourage third parties to act as hidden intermediaries, by collecting individual reports, reporting aggregate beliefs to the principal, and then keeping the gain from arbitrage.

Wide-spread arbitrage can come at a high cost to the principal’s goals. For example, if the principal plans to pool the reports into an aggregate forecast, the result will be distorted.<sup>2</sup> In our example, all colluding agents report the same probability, thereby hiding the true level of disagreement among

<sup>1</sup>For example, the Good Judgment Project (Tetlock and Gardner 2015), the winning competitor in the recent IARPA ACE forecasting competition, employed this technique.

<sup>2</sup>The unweighted arithmetic mean does not change if colluding agents report their group mean, as in our example. However, the principal may wish to compute a weighted mean, a geometric mean (Genest and Zidek 1986), or another aggregate measure such as a supra-Bayesian inference (Morris 1977).

agents, which may cause the principal to have too much confidence in an aggregate forecast. Misreports due to arbitrage also obscure the forecasters' relative accuracy, which is problematic if the principal wants to identify the most accurate forecasters. Fundamentally, colluding agents are extracting additional payment, at the expense of either the principal or some other agents, without contributing correspondingly valuable information.

Chun and Shachter (2011) hoped to develop a mechanism resistant to collusion, but concluded that it "is still an open question whether there is any strictly proper mechanism that does not admit arbitrage, but it seems unlikely." The question remains open. However, if either strict properness or no-arbitrage is slightly weakened, we show that collusion-proof mechanisms do exist.

First, in Section 3, we design a strictly proper mechanism that does not admit arbitrage as long as there is an  $\epsilon > 0$  such that each agent report is between  $\epsilon$  and  $1 - \epsilon$ . In practice, this is a mild restriction. For instance, the prediction market PredictIt<sup>3</sup> (which operates by continuous double auction) does not allow trades at less than 1c or more than 99c, equivalent to setting  $\epsilon = 0.01$ . Systems based on the popular logarithmic scoring rule often impose a similar bound on reports in order to avoid infinite agent losses, which occur whenever an agent assigns probability zero to the outcome that is eventually realized. Each of the mechanisms studied by Chun and Shachter (2011) provides arbitrage possibilities for every possible profile of agent reports with at least some disagreement. In contrast, our mechanism avoids arbitrage in almost all cases, except for agents reporting extremal beliefs. A drawback of our mechanism is that, when choosing  $\epsilon$  small, very large payments may be required. If payments on extreme instances are not high enough, then there exist instances where the average report is close to 0.5 on which all agents receive almost indistinguishable scores. While this may limit the practical viability of our scheme, the existence of a no-arbitrage mechanism for a wide class of agent reports sheds a new optimistic light on Chun and Shachter's conjecture, prompting hope for the existence of a fully arbitrage-free, strictly proper mechanism.

Second, in Section 4, we design weakly proper mechanisms that fully avoid arbitrage for unrestricted agent reports. We begin by studying mechanisms where an agent's payoff depends only on their own report. We fully characterize the set of arbitrage-free and weakly proper scoring rules as those that offer at most two sets of conditional payments, for example paying \$1 if and only if the agent's report is greater than 0.5 for the true outcome. Under these mechanisms, truth-telling is an undominated strategy, but not uniquely. By combining these rules with a rule that pays each agent the score of the median report, we obtain arbitrage-free mechanisms that are *weakly dominant proper*: truth-telling is the unique undominated strategy. Weakly dominant proper mechanisms are close to strictly proper: incentives are not strict only if some profiles are impossible. We consider these weakly dominant proper mechanisms to be compelling practical options for a principal who wants to avoid collusion.

<sup>3</sup><https://www.predictit.org/>

## 2 Preliminaries

Let  $X$  be a Boolean<sup>4</sup> random variable: an *event* with possible *outcomes* 0 and 1. Let  $N$  be a set of  $n$  agents. Each agent has a belief  $p_i \in [0, 1]$  as to the probability that  $X = 1$ , and reports  $\hat{p}_i \in [0, 1]$ . Following the nomenclature of Chun and Shachter (2011), a *contract function*  $\Pi : [0, 1]^n \times \{0, 1\} \rightarrow \mathbb{R}^n$  maps agent reports and the event outcome to payments. A positive payment is a payment from the mechanism to the agent. We refer to agent  $i$ 's payment as  $\Pi_i$ . Contract functions are a general class of mechanisms that include standard proper scoring rules (Brier 1950; Savage 1971; Gneiting and Raftery 2007), market scoring rules (Hanson 2003), competitive scoring rules (Kilgour and Gerchak 2004), and wagering mechanisms (Lambert et al. 2008).

A contract function  $\Pi$  is *weakly proper* if for every agent  $i$  with belief  $p_i$ , every vector of other agents' reports  $\hat{\mathbf{p}}_{-i}$ , and every  $\hat{p}_i$ ,

$$\begin{aligned} & p_i \Pi_i((p_i, \hat{\mathbf{p}}_{-i}), 1) + (1 - p_i) \Pi_i((p_i, \hat{\mathbf{p}}_{-i}), 0) \\ & \geq p_i \Pi_i((\hat{p}_i, \hat{\mathbf{p}}_{-i}), 1) + (1 - p_i) \Pi_i((\hat{p}_i, \hat{\mathbf{p}}_{-i}), 0). \end{aligned}$$

We say that  $\Pi$  is *strictly proper* if the inequality is strict whenever  $\hat{p}_i \neq p_i$ . Thus, for a strictly proper  $\Pi$ , the expected payoff to  $i$  is uniquely maximized when  $i$  reports the true belief  $p_i$ .

A contract function  $\Pi$  *admits arbitrage* if there exists a coalition  $C \subseteq N$  and vectors  $\mathbf{q} = (q_i)_{i \in N}$  and  $\mathbf{r} = (r_i)_{i \in N}$  with  $q_i = r_i$  for all  $i \notin C$  for which

$$\sum_{i \in C} \Pi_i(\mathbf{q}, x) \geq \sum_{i \in C} \Pi_i(\mathbf{r}, x) \quad \text{for each } x \in \{0, 1\},$$

and the inequality is strict for some  $x$ . This definition is adapted from Chun and Shachter (2011), who define a version that requires that arbitrage is possible whenever agents disagree. We say that a contract function that does not admit arbitrage is *arbitrage free*.

A scoring rule,  $s : [0, 1] \times \{0, 1\} \rightarrow \mathbb{R}$ , is a function that assigns a real-valued score to an agent based only on the agent's report and the realized outcome. We say that  $s$  is weakly (resp. strictly) proper if the corresponding contract function  $\Pi$  that pays each agent

$$\Pi_i(\hat{\mathbf{p}}, x) = s(\hat{p}_i, x)$$

is weakly (resp. strictly) proper.

It is easy to see that a contract function is weakly (resp. strictly) proper if, and only if, for all fixed reports of the other agents  $\hat{\mathbf{p}}_{-i}$ , the payment to agent  $i$  takes the form of a weakly (resp. strictly) proper scoring rule.

**Scoring rule characterization.** We will utilize the well-known characterization of proper scoring rules (McCarthy 1956; Savage 1971; Schervish 1989; Gneiting and Raftery 2007).

<sup>4</sup>In the full version we show how to extend our results to any finite discrete outcome space.

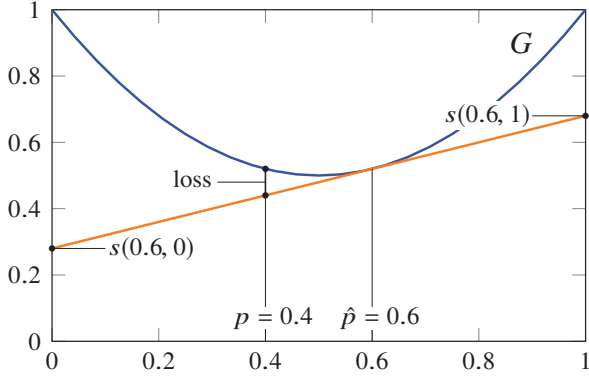


Figure 1: Illustration of Theorem 1 for an agent with belief  $p = 0.4$  who reports  $\hat{p} = 0.6$ . The horizontal axis is belief and the vertical axis is score. The line segment is  $s(0.6; p)$ , the expected score for report 0.6 under belief  $p$ . Its endpoints are the scores  $s(0.6, 0)$  and  $s(0.6, 1)$  in outcomes 0 and 1, respectively.

**Theorem 1.** *Scoring rule  $s$  is (strictly) proper if and only if there exists a (strictly) convex function  $G : [0, 1] \rightarrow \mathbb{R}$  with*

$$s(\hat{p}, x) = G(\hat{p}) + dG(\hat{p}) \cdot (x - \hat{p}),$$

where  $dG(\hat{p})$  is a subgradient of  $G$  at  $\hat{p}$ . Furthermore,  $G(p)$  is the expected score for truthful reporting with belief  $p$ .

Write  $s(\hat{p}; p)$  for the expected score of report  $\hat{p}$  under belief  $p$ . From Theorem 1 and linearity of expectation, we obtain  $s(\hat{p}; p) = G(\hat{p}) + dG(\hat{p}) \cdot (p - \hat{p})$ . Note that  $s(p; p) = G(p)$  as claimed.

A pictorial representation (Gneiting and Raftery 2007) is shown in Figure 1. Given  $G$  and a fixed report  $\hat{p} = 0.6$ , the function  $s(0.6; p)$  traces a *subtangent* line of  $G$ , tangent at  $p = 0.6$ . Its slope is equal to the subgradient  $dG(\hat{p})$ . The agent's score in outcome 0 or 1 is the height of this subtangent at the vertical line  $p = 0$  or  $p = 1$ , respectively. An agent who believes  $p = 0.4$  and reports  $\hat{p} = 0.6$  expects to receive payment equal to the height of the subtangent at the vertical line  $p = 0.4$ , an amount strictly worse than  $G(0.4)$ , their expected payment when reporting truthfully.

### 3 An Arbitrage-Free Rule Under Bounded Reports

In this section we present a strictly proper contract function. It does not permit arbitrage when the reports of the agents are bounded in  $(\epsilon, 1 - \epsilon)$ , for some  $\epsilon > 0$ . We denote our contract function  $M^k$ , where  $k$  is a parameter that is tuned depending on  $\epsilon$ . Payments are defined by  $M_i^k(\hat{\mathbf{p}}, x) = (\sum_{j=1}^n \hat{p}_j - \frac{n}{2})^k + k(x - \hat{p}_i)(\sum_{j=1}^n \hat{p}_j - \frac{n}{2})^{k-1}$ , where  $k$  is an even integer. We first verify that  $M^k$  is strictly proper.<sup>5</sup>

<sup>5</sup>While we may want to bound the reports away from extremes, we do not require any assumptions on the agents' beliefs. If beliefs fall outside of the range of allowable reports, strict properness implies that agents will prefer to report the allowable report closest to their belief.

**Lemma 2.**  $M^k$  is strictly proper.

*Proof.* The result follows from Theorem 1, setting  $G(\hat{p}_i) = (\sum_{j=1}^n \hat{p}_j - \frac{n}{2})^k$ .  $\square$

To gain intuition, let us describe  $M^k$  in words, with the assistance of Figure 2. The blue curve is the function  $\bar{G} : [0, n] \rightarrow \mathbb{R}$  defined by  $\bar{G}(\hat{\mathbf{p}}) = (\sum_{j=1}^n \hat{p}_j - \frac{n}{2})^k$ , which defines payoffs over the space of all possible reports. For fixed reports of other agents  $\hat{\mathbf{p}}_{-i}$  summing to  $c$ , agent  $i$  is faced with a scoring rule derived from  $G(\hat{p}_i) = (c + \hat{p}_i - \frac{n}{2})^k$ , corresponding to a width-1 interval of the function  $\bar{G}$ . An interesting feature of  $M^k$  is that each agent receives the same expected payment from truthful reporting: Pictorially, each agent's payments are determined by a subtangent of  $G$  taken at the same point.

We now show that if it is possible to bound agents' reports away from 0 and 1, then there exists a  $k$  for which  $M^k$  is arbitrage free. The proof proceeds by first showing that the payment a group of agents receives is a function only of the sum of their reports, and that, while all reports are within  $(\epsilon, 1 - \epsilon)$ , this function is increasing for  $X = 1$  and decreasing for  $X = 0$ .

**Theorem 3.** *Suppose there exists an  $\epsilon > 0$  such that  $\hat{p}_j \in (\epsilon, 1 - \epsilon)$  for all agents  $j$ . Then  $M^k$  is arbitrage free whenever  $k > \frac{n+2}{2\epsilon}$ .*

*Proof.* Consider reports  $\hat{\mathbf{p}}$  and a coalition of agents  $C = \{1, \dots, |C|\}$ , where  $|C| \geq 2$ . Let us examine the total payment to agents in  $C$  when  $X = 1$ . For conciseness, given a subset  $S \subseteq N$ , we use the notation  $\hat{p}_S = \sum_{j \in S} \hat{p}_j$  to denote the sum of reports of agents in  $S$ .

$$\begin{aligned} & \sum_{j \in C} M_j^k(\hat{\mathbf{p}}, 1) \\ &= \sum_{j \in C} ((\hat{p}_N - \frac{n}{2})^k + k(1 - \hat{p}_j)(\hat{p}_N - \frac{n}{2})^{k-1}) \\ &= |C|(\hat{p}_N - \frac{n}{2})^{k-1} \left( \hat{p}_N - \frac{n}{2} + \frac{k}{|C|} \sum_{j \in C} (1 - \hat{p}_j) \right) \\ &= |C|(\hat{p}_N - \frac{n}{2})^{k-1} (\hat{p}_{N \setminus C} - \frac{n}{2} + k + (1 - \frac{k}{|C|})\hat{p}_C) \end{aligned}$$

Note that this is a function of  $\hat{p}_C$  and  $\hat{p}_{N \setminus C}$  only. Now we differentiate with respect to  $\hat{p}_C = \sum_{j \in C} \hat{p}_j$ :

$$\begin{aligned} & \frac{d}{d(\hat{p}_C)} \left( \sum_{j \in C} M_j^k(\hat{\mathbf{p}}, 1) \right) \\ &= (k-1)|C|(\hat{p}_N - \frac{n}{2})^{k-2} (\hat{p}_{N \setminus C} - \frac{n}{2} + k + (1 - \frac{k}{|C|})\hat{p}_C) \\ & \quad + (1 - \frac{k}{|C|})|C|(\hat{p}_N - \frac{n}{2})^{k-1} \\ &= |C|(\hat{p}_N - \frac{n}{2})^{k-2} ((k-1)(\hat{p}_{N \setminus C} - \frac{n}{2} + k + (1 - \frac{k}{|C|})\hat{p}_C) \\ & \quad + (1 - \frac{k}{|C|})(\hat{p}_N - \frac{n}{2})) \\ &= k|C|(\hat{p}_N - \frac{n}{2})^{k-2} (k-1 + (1 - \frac{1}{|C|})\hat{p}_{N \setminus C} \\ & \quad + (1 - \frac{k}{|C|})\hat{p}_C - \frac{n}{2}(1 - \frac{1}{|C|})) \end{aligned}$$

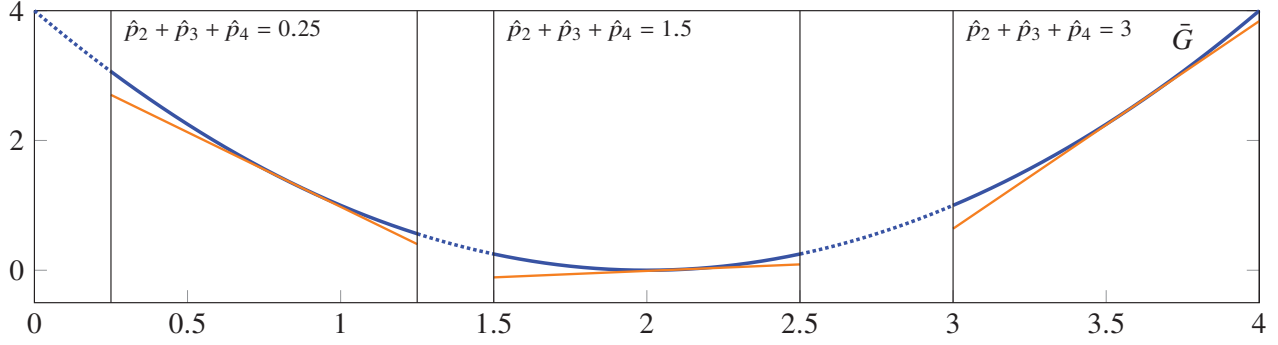


Figure 2: Illustration of contract function  $M^k$ ,  $k = 2$ , for agent 1 of four. The horizontal axis is the sum of all reports  $\hat{p}_i$  and the vertical axis is score. We consider three examples: the sum of the reports of agents 2, 3, and 4 is 0.25 (left), 1.5 (middle), and 3 (right). Each case induces a strictly proper scoring rule for agent 1, as in Figure 1: the expected score function  $\bar{G}$  is shown in solid blue along with subtangent  $s(0.6; p)$  in orange.

We show that the derivative is positive whenever  $k > \frac{n+2}{2\epsilon}$ . The multiplicative factor  $k|C|(\sum_{i=1}^n \hat{p}_i - \frac{n}{2})^{k-2}$  is always positive because  $k$  is even, so we can ignore it and focus on the remaining part of the expression:

$$\begin{aligned} k - 1 + (1 - \frac{1}{|C|})\hat{p}_{N \setminus C} + (1 - \frac{k}{|C|})\hat{p}_C - \frac{n}{2}(1 - \frac{1}{|C|}) \\ > k - 1 - \frac{k}{|C|}|C|(1 - \epsilon) - \frac{n}{2} = k\epsilon - 1 - \frac{n}{2} > 0 \end{aligned}$$

The first inequality comes from removing positive terms and the fact that  $\hat{p}_j < 1 - \epsilon$  for all  $j$ , and the final inequality from the definition of  $k$ .

In particular, the positive derivative implies that any coordination in which the colluding agents decrease the sum of their reports results in them collectively receiving a strictly lower payment when  $X = 1$ . Therefore, any successful arbitrage must have the agents in  $C$  increase the sum of their reports. But, because  $M^k$  is symmetric in the possible outcomes, an identical argument shows that if the agents in  $C$  increase the sum of their reports, their total payment for outcome  $X = 0$  would strictly decrease. Hence, arbitrage is not possible if the colluding agents change the sum of their reports. However, recall that the expression for  $\sum_{j \in C} M_j^k(\hat{\mathbf{p}}, 1)$  is a function of  $\hat{p}_C$  and  $\hat{p}_{N \setminus C}$ . Therefore, any group deviation that does not change the sum  $\hat{p}_C$  does not change the total payment for  $X = 1$  or, symmetrically, for  $X = 0$ , and so cannot constitute arbitrage. This completes the proof.  $\square$

Our next example demonstrates the failure of  $M^k$  to be fully arbitrage free.

**Example 1.** Consider  $M^k$  with  $k = 2$  (for higher  $k$ , a similar example works but requires more extreme reports). Let  $\mathbf{p} = (0, 0, 0, 0.98, 0.98)$ . If  $X = 0$ , each of the last two agents receives payment  $(1.96 - 2.5)^2 - 0.98 \cdot (1.96 - 2.5) = 0.8208$ . If  $X = 1$ , then each of them receives  $(1.96 - 2.5)^2 + 0.02 \cdot (1.96 - 2.5) = 0.2808$ . Suppose they both instead report 0.97 (so  $\hat{\mathbf{p}} = (0, 0, 0, 0.97, 0.97)$ ). Now, each of the last two agents receives payment  $(1.94 - 2.5)^2 - 0.97 \cdot (1.94 - 2.5) = 0.8568$  if  $X = 0$  and  $(1.94 - 2.5)^2 + 0.03 \cdot (1.94 - 2.5) =$

0.2968 if  $X = 1$ . Under the manipulation, agents 4 and 5 receive higher payments in both outcomes, constituting arbitrage.

We end this section by elaborating on the relationship between  $k$ ,  $n$ , and  $\epsilon$ . Even for  $n = 5$  agents and  $\epsilon = 0.1$  (reports bounded between 0.1 and 0.9), our bound on  $k$  requires  $k > 35$ . Since payments are on the order of  $(\sum_{i=1}^n \hat{p}_i - \frac{n}{2})^k$ , the principal may have to make payments as large as  $(5 - \frac{5}{2})^{35} \approx 10^{13}$ . Of course, these payments can be scaled down by any constant while maintaining strict properness, as is standard in elicitation settings to deal with real-world budget constraints, but then it becomes possible that agents in other profiles (say, when  $\sum_{i=1}^n \hat{p}_i \approx \frac{n}{2}$ ) receive payments that are very close to 0, no matter what they report. Designing a rule that provides reasonable incentives to agents while also permitting the principal a reasonable loss in the worst case is an interesting question that we leave open.

## 4 Weakly Proper Arbitrage-Free Rules

In the previous section, we presented a contract function that is strictly proper and arbitrage free when reports are bounded away from 0 and 1. In this section, we relax strict properness to weak properness, with the goal of achieving fully arbitrage-free contract functions.

Weak properness on its own is a very weak guarantee. Paying each agent a constant amount  $\Pi_i(\hat{\mathbf{p}}, x) = d$  is weakly proper, because no agent can profit by reporting  $\hat{p}_i \neq p_i$ . It is also arbitrage free, since the total payment to any group of agents is also fixed by the mechanism in advance. However, this is not a compelling contract function; a constant score can hardly be called a scoring function at all. Agents paid a fixed amount have no incentive to gather or process information if there is even a minimal cost to do so. We will consider two properties to avoid such degenerate contract functions.

First, we want a contract function that distinguishes between good and bad reports. To this end, we say that a contract function  $\Pi$  satisfies *weak distinguishability* if, whenever there exist agents  $i, j$  with  $\hat{p}_i = 0$  and  $\hat{p}_j = 1$ , we have  $\Pi_i(\hat{\mathbf{p}}, 0) > \Pi_j(\hat{\mathbf{p}}, 0)$  and  $\Pi_i(\hat{\mathbf{p}}, 1) < \Pi_j(\hat{\mathbf{p}}, 1)$ . An agent

with a maximally accurate report should receive a strictly higher payment than an agent with a maximally inaccurate report. While this may seem like a trivial property, it turns out that mechanism  $M^k$  from Section 3 actually fails weak distinguishability.

Second, observe that under the constant payment rule, agents can misreport *safely*: even without knowing the reports of the other agents, an agent can misreport and be guaranteed to not regret their misreport once the other reports are revealed. If a contract function is weakly proper and does not permit safe misreports, we say that it is weakly dominant proper. Formally, a weakly proper contract function  $\Pi$  is weakly dominant proper if, for any agent  $i$  with belief  $p_i$  and any report  $\hat{p}_i \neq p_i$ , there exist reports  $\hat{\mathbf{p}}_{-i}$  of the other agents such that

$$\begin{aligned} & p_i \Pi_i((p_i, \hat{\mathbf{p}}_{-i}), 1) + (1 - p_i) \Pi_i((p_i, \hat{\mathbf{p}}_{-i}), 0) \\ & > p_i \Pi_i((\hat{p}_i, \hat{\mathbf{p}}_{-i}), 1) + (1 - p_i) \Pi_i((\hat{p}_i, \hat{\mathbf{p}}_{-i}), 0). \end{aligned}$$

Thus, for any misreport, there is a possible configuration of others' reports such that the expected value of the misreport is strictly lower than the expected value of reporting truthfully. Stated in game-theoretic terms, reporting truthfully is a weakly dominant strategy.<sup>6</sup> Weakly dominant properness was implicitly defined, but not named, by Freeman, Pennock, and Wortman Vaughan (2017).

#### 4.1 Weakly Distinguishing Scoring Rules

Let us more closely examine contract functions that pay all agents according to a fixed scoring rule. Chun and Shachter (2011) showed that all such rules admit arbitrage whenever  $s$  is strictly proper. What if  $s$  is only weakly proper? We know from our discussion above that there exists at least one such rule—the constant payment rule—that is arbitrage free. In this section, we provide a complete characterization of the class of arbitrage-free weakly proper scoring rules.

For any convex function  $G$ , write  $G(q) = \max_{i \in \mathcal{I}} S^i(q)$ , for some index set  $\mathcal{I}$ , where each  $S^i$  is an affine function defining contingent payments  $S_0^i$  and  $S_1^i$ . Reporting  $\hat{p}$  is equivalent to choosing  $S^i$  where  $i = \arg \max_{i \in \mathcal{I}} S^i(\hat{p})$ , the affine function tangent to  $G$  at  $\hat{p}$ . As a technicality, we assume no redundant choices; that is, each  $S^i$  is the unique optimal choice in expectation for at least one belief.

Say that  $G$  consists of  $\ell$  choices if it is piecewise linear with  $\ell \geq 1$  linear pieces, and call the corresponding weakly proper scoring rule  $s$  an  $\ell$ -choice scoring rule.

**Theorem 4.** *Suppose that a contract function pays all agents according to a single proper scoring rule  $s$ . Then the contract function is arbitrage free if and only if  $s$  is a 1-choice or 2-choice scoring rule.*

*Proof.* If  $s$  is 1-choice, then the corresponding contract function is clearly arbitrage free since payments do not depend on reports. If  $s$  is 2-choice, then the report space  $[0, 1]$  is divided

<sup>6</sup>Translated to notions of dominance in game theory (Shoham and Leyton-Brown (2008) provide an introduction), strict properness corresponds to truth-telling being a strictly dominant strategy, weakly dominant properness corresponds to weak dominance, and weak properness corresponds to very weak dominance.

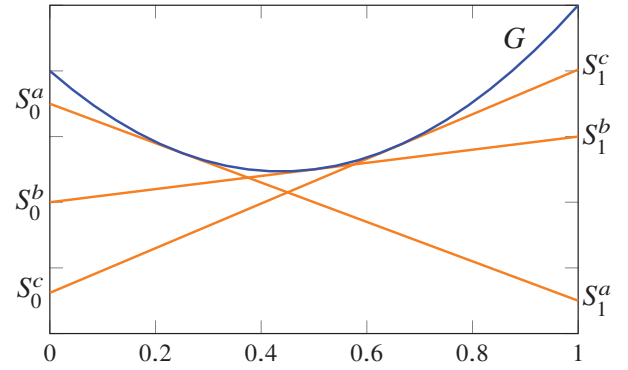


Figure 3: A generic scoring rule that is not 1-choice or 2-choice, with three particular choices highlighted.

into a low and a high region, offering one affine function for each. Call these affine functions  $S^\ell$  and  $S^h$ . By convexity, it must be the case that  $S_0^\ell > S_0^h$  and  $S_1^\ell < S_1^h$ . Any rearrangement of the colluding agents' reports in which more agents receive contingent payments  $S^h$  decreases their total payment if  $X = 0$ , while the reverse is true if the rearrangement results in more agents with contingent payments  $S^\ell$ . If the rearrangement does not change the number of colluding agents receiving contingent payments  $S^h$  and  $S^\ell$ , then the total payment remains unchanged for both  $X = 0$  and  $X = 1$ . Therefore no arbitrage opportunity exists.

For the reverse direction, suppose that  $s$  is not 1-choice or 2-choice. So  $G$  is a maximum over at least three affine functions  $S^a, S^b, S^c$ , each of which is the unique maximizer for at least some  $q \in [0, 1]$ . Generically, the situation must be as pictured in Figure 3: convexity of  $G$  dictates that  $S_0^a > S_0^b > S_0^c$  and  $S_1^c > S_1^b > S_1^a$ . We will show that some fractions of agents with contingent payments  $S^a$  and  $S^c$  can arbitrage by all switching to  $S^b$ . Let  $p$  be a belief for which  $S^b$  is uniquely optimal (note  $0 < p < 1$  by construction); then in particular

$$\begin{aligned} pS_1^b + (1 - p)S_0^b &> pS_1^a + (1 - p)S_0^a \\ pS_1^b + (1 - p)S_0^b &> pS_1^c + (1 - p)S_0^c \end{aligned}$$

implying that, for all  $\alpha \in [0, 1]$ ,

$$\begin{aligned} pS_1^b + (1 - p)S_0^b & > p(\alpha S_1^a + (1 - \alpha)S_1^c) + (1 - p)(\alpha S_0^a + (1 - \alpha)S_0^c) \end{aligned} \quad (1)$$

Suppose an  $\alpha$  fraction of agents are choosing  $S^a$  and a  $1 - \alpha$  fraction choose  $S^c$ . Consider the conditions for which they have a higher average payoff than if they all chose  $S^b$ , under outcomes  $X = 0$  and  $X = 1$  respectively. Because the payoffs are strictly monotone affine functions of  $\alpha$ , there exist unique thresholds  $\bar{\alpha}, \underline{\alpha}$  such that:

$$\alpha S_0^a + (1 - \alpha)S_0^c \geq S_0^b \iff \alpha \geq \bar{\alpha} \quad (2)$$

$$\alpha S_1^a + (1 - \alpha)S_1^c \geq S_1^b \iff \alpha \leq \underline{\alpha} \quad (3)$$

Now we claim  $\underline{\alpha} < \bar{\alpha}$ : if there exists  $\alpha$  for which the inequalities (2) and (3) hold simultaneously, we get a contradiction

with inequality (1). Intuitively, if the mixture of  $S^a$  and  $S^c$  is better on average both when  $X = 0$  and when  $X = 1$ , then the mixture would be better on average for belief  $p$ . This implies  $\underline{\alpha} < \bar{\alpha}$ .

So let  $\alpha^*$  be a rational number satisfying  $\underline{\alpha} < \alpha^* < \bar{\alpha}$ . Write  $\alpha^* = \frac{n}{n+m}$  where  $n, m \in \mathbb{N}$  and construct an instance with  $n$  agents initially choosing  $S^a$  while  $m$  agents initially choose  $S^c$ . Now suppose all agents switch to selecting  $S^b$ . By (2) and (3) and the choice of  $\alpha^*$ , their average payoff strictly improves both when  $X = 0$  and  $X = 1$ . This proves an arbitrage opportunity.  $\square$

If  $s$  is 1-choice then every agent receives the same payment, and so the contract function fails weak distinguishability. However, when  $s$  is 2-choice, agents are separated into two groups based on their reports. Those with higher reports (including  $\hat{p}_i = 1$ ) get paid more than those with lower reports (including  $\hat{p}_i = 0$ ) when  $X = 1$ , and less when  $X = 0$ . Therefore these rules satisfy weak distinguishability. They are not weakly dominant proper since an agent is always indifferent between any two reports that induce the same contingent payments.

As a natural example of a 2-choice scoring rule, consider the 0-1 score paying \$1 iff the agent's report favors the true outcome:  $\hat{p} \geq 0.5$  and  $X = 1$  or  $\hat{p} < 0.5$  and  $X = 0$ . This rule corresponds to the 2-piecewise-linear function  $G(\hat{p}) = \max\{\hat{p}, 1 - \hat{p}\}$ .

## 4.2 Weakly Dominant Properness and the Median Rule

We now define a contract function that is arbitrage free and weakly dominant proper. Suppose for convenience that  $n$  is odd and fix a strictly-proper scoring rule  $s$ . We will simply pay each agent the score (according to  $s$ ) of the *median* report.<sup>7</sup> Formally,

$$\Pi_i(\hat{\mathbf{p}}, x) = s(\text{med}(\hat{\mathbf{p}}), x).$$

The median rule, like any rule that pays according to an aggregate statistic of all reports, is clearly arbitrage free. In order for a group of colluding agents to change the payments made by the mechanism, they must change the value of the aggregate statistic, which strictly decreases the total payment for some outcome.

Additionally, the median mechanism is weakly dominant proper. For weak properness, consider an agent  $i$  with belief  $p_i$  and report  $\hat{p}_i$ . It is easy to see that either  $\text{med}(\hat{\mathbf{p}}) \leq \text{med}((p_i, \hat{\mathbf{p}}_{-i})) \leq p_i$  or  $\text{med}(\hat{\mathbf{p}}) \geq \text{med}((p_i, \hat{\mathbf{p}}_{-i})) \geq p_i$ . In either case,  $i$  achieves a (weakly) higher expected payment by reporting  $p_i$  than  $\hat{p}_i$ .<sup>8</sup>

To show that the median mechanism satisfies the additional requirement for weakly dominant properness, suppose that agent  $i$  reports  $\hat{p}_i > p_i$  (the case  $\hat{p}_i < p_i$  can be handled symmetrically). We must exhibit reports  $\hat{\mathbf{p}}_{-i}$  of the other agents for which  $i$  would have strictly preferred to report  $p_i$ . To do so, suppose that  $\frac{n-1}{2}$  of the other agents report  $\hat{p}_j < p_i$

<sup>7</sup>If  $n$  is even, we can use the left or right median instead.

<sup>8</sup>That an agent's expected score is a single-peaked function of their report follows easily from Theorem 1.

and  $\frac{n-1}{2}$  of them report  $\hat{p}_i < \hat{p}_j$ . If  $i$  had truthfully reported  $p_i$ , the median report would have been her report  $p_i$  and she would receive payment  $s(p_i, x)$ . Instead, the median report is her misreport  $\hat{p}_i$  and she receives payment  $s(\hat{p}_i, x)$ . By strict properness of  $s$ , agent  $i$  strictly prefers the former. We have proven the following theorem.

**Theorem 5.** *The median payment rule is arbitrage free and weakly dominant proper.*

The median payment rule does not satisfy weak distinguishability, because it pays all agents identically.

## 4.3 Weakly Distinguishing and Weakly Dominant Proper Contract Functions

We now combine the mechanisms from each of the previous subsections to obtain a contract function that preserves the desirable properties of both of them. Let  $s$  be a 2-choice proper scoring rule, with corresponding convex function  $G$  consisting of affine functions  $S^\ell$  and  $S^h$ , defining contingent payments  $S_0^h < S_0^\ell$  and  $S_1^h > S_1^\ell$ . Let  $c = \min\{S_0^\ell - S_0^h, S_1^h - S_1^\ell\}$ . Further, let  $r$  be a strictly-proper scoring rule bounded in the interval  $[0, 1]$ .

Define contract function  $s^+$  by

$$s_i^+(\hat{\mathbf{p}}, x) = s(\hat{p}_i, x) + \frac{c}{n+1} \cdot r(\text{med}(\hat{\mathbf{p}}), x).$$

That is,  $s^+$  pays each agent according to the sum of a 2-piecewise-linear proper scoring rule, and some sufficiently small fraction of the median rule. In doing so, it inherits the weak distinguishability of scoring rule  $s$  and the weakly dominant properness of the median rule, while preserving arbitrage-freeness.

**Theorem 6.** *The contract function  $s^+$  is weakly distinguishing, weakly dominant proper, and arbitrage free.*

*Proof.* Weak distinguishability follows immediately from the definition. For weakly dominant properness, we note that  $s^+$  is weakly proper since it is the sum of two weakly proper rules. For any report  $\hat{p}_i \neq p_i$ , there exist reports of the other agents for which  $i$  would be better off reporting  $p_i$  than  $\hat{p}_i$  under the median rule. Therefore,  $i$  would be better off reporting  $p_i$  than  $\hat{p}_i$  under  $s^+$  too.

For arbitrage freeness, consider some colluding coalition  $C$ , and two sets of reports  $\mathbf{q}$  and  $\mathbf{q}'$  with  $q_i = q'_i$  for all  $i \notin C$ . Consider scoring rule  $s$  from the definition of  $s^+$ . Denote by  $L \subset [0, 1]$  the low region of the report space, in which agents receive contingent payments defined by  $S^\ell$ . Let  $n_L = |\{i \in C : q_i \in L\}|$  and  $n'_L = |\{i \in C : q'_i \in L\}|$ . Let  $H = [0, 1] \setminus L$  denote the high region, where agents receive contingent payments defined by  $S^h$ , with  $n_H = |C| - n_L$  and  $n'_H = |C| - n'_L$ . We consider two cases.

**Case 1:** Suppose that  $n_L = n'_L$ , which implies  $n_H = n'_H$ .

Then for each  $x \in \{0, 1\}$  we have

$$\begin{aligned}
& \sum_{i \in C} s_i^+(\mathbf{q}, x) \\
&= n_L S_x^\ell + n_H S_x^h + |C| \frac{c}{n+1} r(\text{med}(\mathbf{q}), x) \\
&= n'_L S_x^\ell + n'_H S_x^h + |C| \frac{c}{n+1} r(\text{med}(\mathbf{q}), x) \\
&= \sum_{i \in C} s_i^+(\mathbf{q}', x) + |C| \frac{c}{n+1} (r(\text{med}(\mathbf{q}), x) - r(\text{med}(\mathbf{q}'), x))
\end{aligned}$$

By properness of  $r$ , if  $r(\text{med}(\mathbf{q}), x) > r(\text{med}(\mathbf{q}'), x)$  for some  $x \in \{0, 1\}$ , then it must be the case that  $r(\text{med}(\mathbf{q}), 1 - x) < r(\text{med}(\mathbf{q}'), 1 - x)$ . Therefore, no arbitrage opportunity exists.

**Case 2:** Suppose that  $n_L - n'_L = n'_H - n_H > 0$ . The opposite case is symmetric. Then we have

$$\begin{aligned}
& \sum_{i \in C} s_i^+(\mathbf{q}, 0) \\
&= n_L S_0^\ell + n_H S_0^h + |C| \frac{c}{n+1} r(\text{med}(\mathbf{q}), 0) \\
&= n'_L S_0^\ell + n'_H S_0^h + (n_L - n'_L)(S_0^\ell - S_0^h) \\
&\quad + |C| \frac{c}{n+1} r(\text{med}(\mathbf{q}), 0) \\
&\geq n'_L S_0^\ell + n'_H S_0^h + c + |C| \frac{c}{n+1} r(\text{med}(\mathbf{q}), 0) \\
&\geq n'_L S_0^\ell + n'_H S_0^h + c + |C| \frac{c}{n+1} (r(\text{med}(\mathbf{q}'), 0) - 1) \\
&> n'_L S_0^\ell + n'_H S_0^h + |C| \frac{c}{n+1} r(\text{med}(\mathbf{q}'), 0) = \sum_{i \in C} s_i^+(\mathbf{q}', 0)
\end{aligned}$$

The first inequality holds because  $n_L - n'_L \geq 1$  and  $S_0^\ell - S_0^h \geq c$ , the second inequality because  $r$  is bounded in  $[0, 1]$ , and the final inequality because  $|C| < n + 1$ .

That  $\sum_{i \in C} s_i^+(\mathbf{q}, 1) < \sum_{i \in C} s_i^+(\mathbf{q}', 1)$  follows similarly, implying that no arbitrage opportunity exists.  $\square$

Note that arbitrage freeness does not in general follow from combining two arbitrage-free rules, and that the subtlety in the definition of  $s^+$  is necessary. Suppose that we had instead defined  $s^+$  without a small enough scaling of the median rule. In particular, suppose that  $s$  is the 0-1 score defined earlier and  $r$  is the quadratic score ( $s(\hat{p}_i, x) = 1 - (\hat{p}_i - x)^2$ ). Consider the contract function that pays each agent  $s(\hat{p}_i, x) + r(\text{med}(\hat{\mathbf{p}}), x)$ . If  $\mathbf{p} = (1, 1, 0)$  then the total payment to all agents is \$1 if  $X = 0$  and \$5 if  $X = 1$ . But if instead the agents report  $\hat{\mathbf{p}} = (0.5, 0.5, 0.5)$ , their total payment is \$2.25 if  $X = 0$  and \$5.25 if  $X = 1$ .

## 5 Conclusion

We have explored mechanisms for truthfully eliciting probabilistic forecasts from individuals who may collude with one another. When strict incentives for truth-telling are required, the principal can avoid arbitrage provided that reports are bounded away from extreme probabilities. If incentives can be relaxed, then arbitrage can be avoided with a mechanism in which truth-telling remains the unique undominated strategy.

So far, we have only considered binary events. In the full version, we provide an inductive construction that extends contract functions defined for a binary random variable to

any random variable  $X$  with finitely many values. The constructed contract function first pays agents based on their prediction for the event  $X = 0$ , and, if the event  $X \neq 0$  occurs, pays agents based on their predictions conditioned on the event  $X \neq 0$ . Our construction preserves properness of various strengths and arbitrage freeness. The mechanisms we have presented can therefore be extended in a straightforward manner to elicit predictions for any discrete random variable.

Compelling and challenging open questions remain. In particular, we still do not know whether a strictly proper and fully arbitrage-free contract function exists. In Section 3, we discussed an easier open problem: Can a mechanism with similar guarantees to  $M^k$  be designed that has stronger truth-telling incentives in practice? Towards an impossibility result, one could imagine adding additional constraints to the mechanism design problem such as budget balance, requiring the mechanism to pay out the same amount in total regardless of the reports.

**Acknowledgments.** Dominik Peters was supported by ERC under grant 639945 (ACCORD). Work conducted while Bo Wagoner was at Microsoft Research.

## References

- Brier, G. 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 78:1–3.
- Chun, S., and Shachter, R. D. 2011. Strictly proper mechanisms with cooperating players. In *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence (UAI)*, 125–134.
- Freeman, R.; Pennock, D. M.; and Wortman Vaughan, J. 2017. The double clinching auction for wagering. In *Proceedings of the 18th ACM Conference on Economics and Computation (EC)*, 43–60.
- French, S. 1985. Group consensus probability distributions: A critical survey. *Bayesian statistics 2*.
- Genest, C., and Zidek, J. V. 1986. Combining probability distributions: A critique and an annotated bibliography. *Statistical Science* 1(1):114–148.
- Gneiting, T., and Raftery, A. E. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102(477):359–378.
- Hanson, R. 2003. Combinatorial information market design. *Information Systems Frontiers* 5(1):107–119.
- Kilgour, D. M., and Gerchak, Y. 2004. Elicitation of probabilities using competitive scoring rules. *Decision Analysis* 1(2):108–113.
- Lambert, N. S.; Langford, J.; Wortman, J.; Chen, Y.; Reeves, D. M.; Shoham, Y.; and Pennock, D. M. 2008. Self-financed wagering mechanisms for forecasting. In *Proceedings of the 9th ACM Conference on Electronic Commerce (EC)*, 170–179.
- McCarthy, J. 1956. Measures of the value of information. *Proceedings of the National Academy of Sciences* 42(9):654–655.

- Morris, P. A. 1977. Combining expert judgments: A Bayesian approach. *Management Science* 23(7):679–693.
- Savage, L. J. 1971. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association* 66(336):783–801.
- Schervish, M. J. 1989. A general method for comparing probability assessors. *The Annals of Statistics* 17(4):1856–1879.
- Shoham, Y., and Leyton-Brown, K. 2008. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press.
- Tetlock, P. E., and Gardner, D. 2015. *Superforecasting: The Art and Science of Prediction*. New York, NY, USA: Crown Publishing Group.