Bounded Rationality in Wagering Mechanisms

David M. Pennock
Microsoft Research

Vasilis Syrgkanis
Microsoft Research

Jennifer Wortman Vaughan
Microsoft Research

Abstract

Wagering mechanisms allow decision makers to inexpensively collect forecasts from groups of experts who reveal their information via bets with one another. Such mechanisms naturally induce a game in which strategic considerations come into play. What happens in the game depends on the reasoning power of the experts. At one extreme, if experts are fully rational, no-trade theorems imply no participation. At the other extreme, if experts ignore strategic considerations, even the least informed will wager as if his beliefs are correct. Economists have analyzed the former case and decision theorists the latter, but both are arguably unrealistic. In this paper, we adopt an intermediate model of bounded rationality in wagering mechanisms based on level-k reasoning. Under this model, overconfidence allows some participation to be sustained, but experts who realize they are at a relative disadvantage do bow out. We derive conditions on the particular wagering mechanism used under which participation is unbiased, and show that unbiasedness always implies truthful reports. We show that if participation is unbiased, then participation rates unavoidably fall as players' rationality increases, vanishing for large k. Finally, we zoom in on one particular information structure to give a complete characterization specifying the conditions under which mechanisms are unbiased and show how to maximize participation rates among all unbiased mechanisms.

1 INTRODUCTION

A wagering mechanism is an inexpensive tool to elicit forecasts from a group. Rather than paying for the information directly, a decision maker can let members of the group wager with each other and, in the process, capture their beliefs. In the ideal case, everyone with information has incentive to reveal their true subjective forecasts without bias at little or no cost to the decision maker.

Kilgour and Gerchak (2004) proposed one such wagering mechanism, called a shared scoring rule. Theoretically, this mechanism is individually rational (players prefer participating over not participating), truthful, and budget balanced, meaning the decision maker can collect honest, accurate forecasts from every member of the group and pay nothing. However, individual rationality and truthfulness rely on the key assumption that players have *immutable beliefs* (Lambert et al., 2008; Chun and Shachter, 2011). That is, players believe what they believe and do not update their beliefs even when matched against opponents. They are oblivious of the fact that they are playing against reasoning agents in a zero-sum game. They employ what might be termed *level-0 reasoning*.

Immutable beliefs yield an inherent contradiction. The sum of players' private expected profits is positive: everyone who agrees to play expects to gain. Yet the true sum of their profits is zero. Everyone therefore knows that at least one player must be overly optimistic. The immutable beliefs assumption may be defensible for a one-shot, isolated game, but in an iterative game, at least one player should rapidly observe a disagreement between what he expects and what he receives, making it reasonable to assume that he will adapt and update over time. In an iterative mental game prior to play, updating may occur ex ante. All players know that someone's information must be inaccurate, so it is natural to imagine revisions even in a one-shot game.

Game theory predicts nearly the opposite behavior in a wagering mechanism. Under the relatively weak assumption of a common prior, if all players are rational, know that all players are rational, and know that all players know that all players are rational, ad infinitum, then speculative wagers should not happen at all (Milgrom and Stokey, 1982). The decision maker would not get a single report from any member of the group. Intuitively, every proffered wager would be declined as evidence of superior information. Private expectations could not differ from outcomes in a systematic way, and an overall bias toward optimism could not be sustained in a rational equilibrium (Lucas, 1972; Nielsen et al., 1990).

To design a wagering mechanism that encourages relatively broad, unbiased, and truthful participation, we need to understand how people behave. Do they act with immutable beliefs? Do they reason with complete rationality? Or do they fall somewhere in between?

The implications of unbounded rationality are absurd. If true, players could never "agree to disagree" (Aumann, 1976; Geanakoplos and Polemarchakis, 1982) and would never place any wager or make any investment based on a difference of opinion. Refuting this is the simple fact that speculative trade happens in exchanges around the world during every second of every day. Transactions do occur between overconfident zero-sum opponents who both expect to gain (Dubey et al., 1987; Jordan and Radner, 1979).

Yet ignoring game theory altogether is an idealization too. People are stubborn, but not entirely naive (Plott and Sunder, 1988). An opponent eager to make a large bet should give anyone pause. (Did I miss something? Is there something my opponent knows that I don't?) Only the most unsophisticated player would ignore the inconsistency between individual and aggregate expectations or repeatedly ignore a systematic difference between his private belief and his experience.

In this paper, we adopt what we view as a more realistic model of players and examine its implications for the design of wagering mechanisms. Our players are boundedly rational: they are neither superhuman *level-*∞ reasoners nor oblivious level-0 reasoners. Instead, building on a vast literature from behavioral game theory, we make the increasingly common assumption that each player reasons at an intermediate level k, treating all of her opponents as level-$(k-1)$ reasoners, with level-0 forming the base of the induction. When $k > 1$, players recognize that they are playing a game against strategic opponents. Still, their reasoning is incomplete. They retain a form of overconfidence in that each player believes that she is one level more capable than her opponents. We find that even this small dose of overconfidence is enough to induce participation.

To obtain accurate information from wagers, we seek shared scoring rules that encourage high rates of participation, unbiased participation (meaning that players don't decide whether to opt in or out based on the direction of their signals), and truthful participation. We first show that under very general assumptions, if an instantiation of the Kilgour-Gerchak mechanism is unbiased, it automatically leads level-k players to report their beliefs truthfully conditioned on participating. We next give a general characterization of which players choose to participate at each level. Roughly speaking, at low levels of rationality (small k) participation rates can be high due to widespread overconfidence, while at high levels of rationality (large k), only the players with the most accurate information choose to participate. We show that for any unbiased instantiation of Kilgour-Gerchak, participation rates shrink to zero as the level of rationality of players grows, illustrating that while unbiasedness is in some ways desirable, it comes at a cost.

The question of how to design unbiased mechanisms does not have a clean analytical answer in the general case. To gain intuition, we therefore zoom in on a particular symmetric information structure that permits tractable analysis under the level-k model. For this information structure, we give a complete characterization specifying the conditions under which the Kilgour-Gerchak mechanism leads to unbiased participation. Interestingly, we find that among all instantiations of Kilgour-Gerchak that are unbiased, the ones that lead to the highest level of participation are those in which players have the smallest incentive to report their true beliefs as opposed to any other forecast as they are rewarded similarly either way. This makes intuitive sense; since a player can only profit if other players lose, rewarding players more evenly has the effect of scaring off fewer of those who are relatively less informed, leading to higher overall participation.

We conclude with a brief discussion of how our work fits into the larger agenda of *behavioral mechanism design* (Ghosh and Kleinberg, 2014; Easley and Ghosh, 2015).

1.1 RELATED WORK

Over the past few decades, several theories have emerged to explain the inconsistencies that often arise between subjects' observed behavior in both lab and field studies and their predicted equilibrium behavior. Brocas et al. (2014) divide these theories into two categories. Theories of *imperfect choice*, such as quantal response equilibrium (McKelvey and Palfrey, 1995), assume that players fully analyze all available information but make noisy decisions or assume that other players make noisy decisions. Theories of *imperfect attention*, such as level-k, assume that players do not fully analyze all available information and therefore make imperfect choices compared with fully rational agents. In this paper we focus on the latter as such theories provide a middle ground between the extreme assumptions of immutable beliefs and full rationality under which one-shot wagering mechanisms have been analyzed in the past.

The earliest proponents of the level-k model were Stahl and Wilson (1994, 1995) and Nagel (1995). The theory was further developed, modified, and empirically evaluated by many others, including Ho et al. (1998), Costa-Gomes et al. (2001), Bosch-Domènech et al. (2002), Costa-Gomes and Crawford (2006), Crawford and Iriberri (2007), and Shapiro et al. (2014). Camerer et al. (2004) introduced a variant of the level-k model, the *cognitive hierarchy model*, in which a player at level k believes that other players' levels are sampled according to some distribution over levels $k' < k$. In particular, the model assumes that there is a Poisson distribution with parameter τ over all players' true levels, and a player at level k believes that the levels of his opponents are distributed according to a normalized Poisson over levels strictly less than k; that is, his beliefs about the relative frequencies of lower levels are correct, but he incorrectly assumes that no other players are capable of reasoning that is at least as sophisticated as his own. As Brocas et al. (2014) point out, the cognitive hierarchy model is typically more difficult to work with analytically than the basic level-k model, often failing to yield crisp and testable predictions of behavior, and has the additional parameter τ to contend with. Additionally, Wright and Leyton-Brown (2010) found that its predictive performance is similar to the basic level-k model on a variety of data sets from the behavioral game theory literature. For these reasons, we primarily focus on the basic level-k model in our analysis, though some of our results could be extended to hold under the cognitive hierarchy model.

There have been several experimental studies examining behavior in one-shot betting games with private information. In early work aimed at testing the predictions of no-trade theorems in the lab, Sonsino et al. (2001) designed a betting game in which fully rational agents would choose not to participate and found substantial rates of betting among subjects. Sløvik (2009) later replicated these results. Rogers et al. (2009) used the same betting game in a new set of experiments in order to compare how well the quantal response model, cognitive hierarchy model, and various hybrids fit the behavior of subjects in the lab. They found evidence of both imperfect choice and imperfect attention in their subjects; the cognitive hierarchy and quantal response models fit the data equally well. Finally, Brocas et al. (2014) used mouse-tracking software in order to gain a better understanding of the cognitive processes behind subjects' actions in a betting game. In their clever design, payoffs of the bet were hidden in opaque boxes. Subjects could view their own payoffs or their opponents' payoffs in different states of the world only by clicking on the appropriate box. By keeping track of which payoffs subjects viewed, the authors were able to make inferences about how many levels of reasoning they were performing. They found a reasonable fit between subjects' actions and the basic level-k model, with different clusters of subjects behaving similarly to what would be expected of level-1, level-2, and level-3 players, though they observed some evidence of imperfect choice as well.

2 PRELIMINARIES AND MODEL

We begin with a review of strictly proper scoring rules and the Kilgour-Gerchak mechanism. We then present our general model of incomplete information and player beliefs and review the level-k model of behavior.

2.1 THE KILGOUR-GERCHAK MECHANISM

We consider a simple scenario in which a set $N = \{1, \cdots, n\}$ of players wager on the value of an unknown binary random variable $X \in \{0, 1\}$. This random variable could represent, for example, the winner of an election, whether or not a product ships on time, or the outcome of a game. We use x to denote a realization of X.

The wagering mechanisms that we analyze are those of Kilgour and Gerchak (2004). All players simultaneously choose whether or not to make a wager. If player i participates, he wagers \$1 and reports a probability $\hat{p}_i$ that $X = 1$. Next, the true state x is revealed, and each player who chose to participate receives a payment that depends on x and the reports of all participating players. Payments are designed to be budget-balanced, meaning that the mechanism's operator takes on no risk. They are also truthful and individually rational for risk-neutral players with immutable beliefs. This means that such players maximize their expected utility by choosing to participate and reporting their true beliefs about the likelihood that $X = 1$.

To achieve truthfulness, Kilgour-Gerchak mechanisms build on the extensive literature on *proper scoring rules* (Savage, 1971; Gneiting and Raftery, 2007). A proper scoring rule is a reward function designed to elicit truthful predictions from risk-neutral agents. A scoring rule S mapping a probability $q \in [0, 1]$ and outcome $x \in \{0, 1\}$ to a real-valued score is *proper* if for all $p \in [0, 1]$, if $X = 1$ with probability p, then the quantity

$$\mathbb{E}[S(q, X)] = p \cdot S(q, 1) + (1 - p) \cdot S(q, 0)$$

is maximized at $q = p$. It is *strictly proper* if this maximum is unique. A common example of a strictly proper scoring rule is the Brier score (Brier, 1950), defined as

$$S(q, x) = 1 - (q - x)^2.$$

Note that the Brier score is bounded in $[0, 1]$. Any bounded scoring rule can be renormalized to lie in this range.

We make heavy use of the following characterization of proper scoring rules from Gneiting and Raftery (2007).

Theorem 1 (Gneiting and Raftery (2007)) *A scoring rule S is (strictly) proper if and only if there exists a (strictly) convex function G, referred to as the* entropy function*, such that for all $q \in [0, 1]$ and all $x \in \{0, 1\}$,*

$$S(q, x) = G(q) + G'(q)(x - q),$$

where G' is any subderivative of G. Moreover, if $X = 1$ with probability p, then $E[S(p, X)] = G(p)$.

The Kilgour-Gerchak mechanism rewards each player by comparing his score to the average score of all other participants. Let $\mathcal{P} \subseteq N$ be the set of players that choose to participate; $\mathcal{P}$ is a random variable that depends on the model of trader behavior. Let $\mathcal{P}_{-i}$ denote all members of $\mathcal{P}$ except i. (We use similar set notation throughout.) The mechanism is defined as follows.

Definition 1 (Kilgour-Gerchak mechanism) *Fix a strictly proper scoring rule S bounded in $[0,1]$. All players $i \in \mathcal{P}$ simultaneously report a probability $\hat{p}_i$. If $|\mathcal{P}| \geq 2$, then when x is revealed, the mechanism assigns to each player $i \in \mathcal{P}$ a net profit (payment minus \$1 wager) of*

$$S(\hat{p}_i, x) - \frac{1}{|\mathcal{P}_{-i}|} \sum_{j \in \mathcal{P}_{-i}} S(\hat{p}_j, x).$$

If $|\mathcal{P}| = 1$, it returns the \$1 wager to the single participating player who receives a net profit of 0.

We assume that players are *risk-neutral*, so if player i chooses to participate, then his utility is

$$u_i(\hat{\mathbf{p}}, x) = \begin{cases} S(\hat{p}_i, x) - \frac{\sum_{j \in \mathcal{P}_{-i}} S(\hat{p}_j, x)}{|\mathcal{P}_{-i}|} & \text{if } |\mathcal{P}| > 1, \\ 0 & \text{otherwise.} \end{cases}$$

Since we analyze only the Kilgour-Gerchak mechanism, the design space we consider is the space of all strictly proper scoring rules bounded in $[0,1]$. Selecting the scoring rule S (or equivalently, selecting the corresponding entropy function G) fully defines the mechanism.

2.2 PLAYER BELIEFS AND BEHAVIOR

To discuss level-k behavior, we first need to define the beliefs of players. We begin by considering a general model of incomplete information based on the well-studied model of Aumann (1976). In later sections, we analyze a specific special case of this model.

We imagine a process in which Nature first draws the value of the random variable $X \in \{0,1\}$ and then, conditioned on this value, draws (possibly correlated) random signals $\Sigma_i \in \{1, \cdots, m_i\}$ for each player i. We define a state of the world $\omega = (x, \sigma_1, \cdots, \sigma_n)$ as an outcome x paired with an assignment of signals σ_i to each player i. Let Ω be the set of all mutually exclusive and exhaustive states of the world. Players share a common prior over Ω.

Under the level-k model, the behavior of each player i is characterized by his *level of rationality*. Under the most simple version of the model, a player at some level $k \in \{1, 2, \cdots\}$ assumes that every other player is at level $k-1$ and best responds to the (distribution over) actions such players would take. This can be viewed as a form of overconfidence; every player believes he is slightly "more rational" than everyone else. We define level-0 players to be risk-neutral with immutable beliefs. Such players always participate (as participation is rational under the immutable beliefs assumption (Lambert et al., 2008)) and truthfully bid their posterior beliefs conditioned on their signals. We could have alternatively defined level-0 players to be noise traders, choosing reports at random, as is common in the level-k literature. This definition would then give rise to immutable belief behavior at level 1, and would therefore not change the nature of our results; it would amount to no more than a simple renumbering of levels.

The behavior of a player at a level k consists of two decisions: whether or not to participate, and his report. We denote with $z_i^{(k)}(\sigma_i)$ an indicator variable that is 1 if player i would choose to participate at level k with signal σ_i and 0 otherwise. We denote with $\hat{p}_i^{(k)}(\sigma_i)$ the report of player i at level k with signal σ_i conditioned on participating. Let $\mathcal{P}^{(k)}$ denote the set of players who would participate *if* they were following level-k behavior; this is a random variable since it depends on the realized signals of each player. The functions $z_i^{(k)}$ and $\hat{p}_i^{(k)}$ are valid level-k behaviors if they maximize a player's utility under the assumption that every other player is of level $k-1$. More formally, let

$$\begin{aligned} &U_i^{(k)}(\hat{p}_i, \sigma_i) \\ &\quad = \mathbb{E}\Bigg[\Bigg(S(\hat{p}_i, X) - \frac{\sum_{j \in \mathcal{P}_{-i}^{(k-1)}} S(\hat{p}_j^{(k-1)}(\Sigma_j), X)}{|\mathcal{P}_{-i}^{(k-1)}|}\Bigg) \\ &\qquad\qquad \cdot \mathbb{1}\{\mathcal{P}_{-i}^{(k-1)} \neq \emptyset\} \;\Bigg|\; \Sigma_i = \sigma_i\Bigg] \end{aligned}$$

be the expected utility of player i at level k if he participates and reports $\hat{p}_i$. Then we must have

$$\begin{aligned} \hat{p}_i^{(k)}(\sigma_i) &\in \arg\max_{\hat{p}_i \in [0,1]} U_i^{(k)}(\hat{p}_i, \sigma_i), \\ z^{(k)}(\sigma_i) &= \mathbb{1}\left\{U_i\left(\hat{p}_i^{(k)}(\sigma_i)\right) > 0\right\}. \end{aligned}$$

Note that we assume players participate only if their expected utility is strictly positive and do not participate if their utility is 0. This is consistent with the scoring rule literature in which it is often assumed that players require strict incentives to truthfully report beliefs.

3 A GENERAL CHARACTERIZATION OF LEVEL-k BEHAVIOR

We are broadly interested in understanding when and how wagering mechanisms can be used to elicit accurate information from players in the level-k model of rationality. With our focus limited to Kilgour-Gerchak mechanisms, we can rephrase this question as asking which scoring rules S lead to high levels of participation and accurate reports.

Two crucial notions in our characterization are *unbiased participation* and of *truthful behavior*. Unbiased participation simply requires that a player's choice of whether or not to participate is independent of his signal (and therefore also independent of the outcome X). This is desirable because biased participation could lead to a biased collection of reports, leading in turn to a biased aggregate forecast for the decision maker.

Definition 2 (Unbiased participation) *We say that behavior at a level $k \in \{0, 1, \cdots\}$ is unbiased if every player's decision of whether or not to participate is independent of the value of his signal, i.e., if $\forall i \in N$, $\forall \sigma_i, \sigma_i' \in \{1, \cdots, m_i\}$, $z_i^{(k)}(\sigma_i) = z_i^{(k)}(\sigma_i')$.*

Using the common prior over Ω, we can define the posterior belief of player i about the likelihood of X after observing the signal $\Sigma_i = \sigma_i$ as

$$p_i(\sigma_i) \equiv \Pr[X = 1 | \Sigma_i = \sigma_i]. \quad (1)$$

Players behave truthfully if they report their true posteriors.

Definition 3 (Truthful behavior) *We say that behavior at a level $k \in \{0, 1, \cdots\}$ is truthful if for each player who chooses to participate at level k, reporting his posterior belief conditioned only on his own signal uniquely maximizes his utility, i.e., if $\forall i \in N$, $\forall \sigma_i \in \{1, \cdots, m_i\}$, $p_i(\sigma_i) = \arg\max_{\hat{p}_i \in [0,1]} U_i^{(k)}(\hat{p}_i, \sigma_i)$.*

3.1 UNBIASEDNESS IMPLIES TRUTHFULNESS

With these two definitions in hand, we can prove several basic characterizations of level-k behavior in Kilgour-Gerchak mechanisms. The first shows that unbiased participation automatically leads to truthfulness, providing another argument that unbiased participation is desirable.

Theorem 2 *In the Kilgour-Gerchak mechanism with strictly proper scoring rule S, if behavior at each level $k' < k$ is unbiased, then level-k behavior is truthful.*

Proof: By the assumption of unbiased participation, for any level $k' < k$, the decision of a level k' player of whether or not to participate is independent of his signal, and therefore independent of the true state of the world X. This implies that for any i, the set $\mathcal{P}_{-i}^{(k')}$ is independent of X, and in fact, deterministic.

If $\mathcal{P}_{-i}^{(k-1)} = \emptyset$ then any wager of player i would yield expected utility 0, so player i would not participate. Suppose that $\mathcal{P}_{-i}^{(k-1)} \neq \emptyset$. Then

$$\begin{aligned} U_i^{(k)}(\hat{p}_i, \sigma_i) &= \mathbb{E}\left[S(\hat{p}_i, X) \mid \Sigma_i = \sigma_i\right] \\ &- \frac{1}{|\mathcal{P}_{-i}^{(k-1)}|} \sum_{j \in \mathcal{P}_{-i}^{(k-1)}} \mathbb{E}\left[S(\hat{p}_j^{(k-1)}(\Sigma_j), X) \mid \Sigma_i = \sigma_i\right]. \end{aligned}$$

The second term is independent of the player's report $\hat{p}_i$. Thus the player will behave truthfully if and only if doing so maximizes the first term, i.e., if and only if

$$p_i(\sigma_i) = \arg\max_{\hat{p}_i \in [0,1]} \mathbb{E}\left[S(\hat{p}_i, X) \mid \Sigma_i = \sigma_i\right],$$

which must hold by definition of $p_i(\sigma_i)$ and the fact that S is a strictly proper scoring rule. ■

Thus to design a truthful mechanism, it is sufficient to design a mechanism that encourages unbiased participation.

3.2 CHARACTERIZING PARTICIPATION

Our next few characterization results examine the conditions under which players choose to participate in the wagering mechanism when the mechanism is unbiased. Lemma 1 examines participation at level k when the mechanism is unbiased at all levels $k' < k$. This lemma will prove useful later when we wish to show that a mechanism is unbiased at all levels for specific signal structures.

Lemma 1 *In the Kilgour-Gerchak mechanism with strictly proper scoring rule S, if behavior at each level $k' < k$ is unbiased, then a player i at level k with signal σ_i participates if and only if $|\mathcal{P}_{-i}^{(k-1)}| > 0$ and*

$$\begin{aligned} &\mathbb{E}[S(p_i(\sigma_i), X) | \Sigma_i = \sigma_i] \\ &\quad > \frac{1}{|\mathcal{P}_{-i}^{(k-1)}|} \sum_{j \in \mathcal{P}_{-i}^{(k-1)}} \mathbb{E}[S(p_j(\Sigma_j), X) | \Sigma_i = \sigma_i]. \quad (2) \end{aligned}$$

Proof: Consider any player $i \in N$ at level k. This player would never participate if $|\mathcal{P}_{-i}^{(k-1)}| = 0$ since his expected utility would be 0, so assume that $|\mathcal{P}_{-i}^{(k-1)}| > 0$. By Theorem 2, since participation is unbiased at every level $k' < k$, players at level k and at level $k-1$ behave truthfully. Using this and the form of player utilities immediately yields the lemma. ■

Theorem 3 builds on the previous lemma to show that when participation is unbiased at levels $k' \leq k$, a player at level k chooses to participate if and only if his a priori expected score is higher than the average a priori expected score of all other players who participate at level $k - 1$.

Theorem 3 *In the Kilgour-Gerchak mechanism with strictly proper scoring rule S with associated entropy G, if behavior at each level $k' \leq k$ is unbiased, then player i at level k participates if and only if $|\mathcal{P}_{-i}^{(k-1)}| > 0$ and*

$$\mathbb{E}[G(p_i(\Sigma_i))] > \frac{1}{|\mathcal{P}_{-i}^{(k-1)}|} \sum_{j \in \mathcal{P}_{-i}^{(k-1)}} \mathbb{E}[G(p_j(\Sigma_j))].$$

Proof: Consider any $i \in N$ at level k. Since this player would never participate if $|\mathcal{P}_{-i}^{(k-1)}| = 0$, assume that $|\mathcal{P}_{-i}^{(k-1)}| > 0$. Since we have assumed that participation is unbiased at level k (as well as lower levels), we know that i either chooses to participate at level k regardless of his signal, or chooses not to participate at level k regardless of his signal. If he chooses to participate, then by Lemma 1, Equation (2) holds for all $\sigma_i \in \{1, \cdots, m_i\}$ and therefore

$$\sum_{\sigma_i=1}^{m_i} \Pr[\Sigma_i = \sigma_i]\mathbb{E}[S(p_i(\Sigma_i), X)|\Sigma_i = \sigma_i] > \sum_{\sigma_i=1}^{m_i} \frac{\Pr[\Sigma_i = \sigma_i]}{|\mathcal{P}_{-i}^{(k-1)}|} \sum_{j \in \mathcal{P}_{-i}^{(k-1)}} \mathbb{E}[S(p_j(\Sigma_j), X)|\Sigma_i = \sigma_i]$$

implying that

$$\mathbb{E}[S(p_i(\Sigma_i), X)] > \frac{1}{|\mathcal{P}_{-i}^{(k-1)}|} \sum_{j \in \mathcal{P}_{-i}^{(k-1)}} \mathbb{E}[S(p_j(\Sigma_j), X)].$$

Similarly, if he chooses not to participate then Lemma 1 implies that

$$\mathbb{E}[S(p_i(\Sigma_i), X)] \leq \frac{1}{|\mathcal{P}_{-i}^{(k-1)}|} \sum_{j \in \mathcal{P}_{-i}^{(k-1)}} \mathbb{E}[S(p_j(\Sigma_j), X)].$$

The proof is completed by observing that by Theorem 1 for any $i \in N$,

$$\begin{aligned}\mathbb{E}[S(p_i(\Sigma_i), X)] &= \mathbb{E}_{\Sigma_i}\left[\mathbb{E}_X[S(p_i(\Sigma_i), X)|\Sigma_i]\right] \\ &= \mathbb{E}\left[G(p(\Sigma_i))\right].\end{aligned}$$

■

This theorem implies that if the mechanism is unbiased at all levels, then at every level k, at least the player with the smallest a priori expected score among those who participate at level $k-1$ stops participating. Additionally, no player who stops participating at some level k ever participates at a higher level, since the average score of other (fictitious) participating players is an increasing function of k. Therefore participation goes to zero as the level of rationality grows, coinciding with fully rational behavior. This illustrates that while unbiasedness is in some ways desirable, it also has its costs.

Corollary 1 *In the Kilgour-Gerchak mechanism with strictly proper scoring rule S, if behavior at every level is unbiased, then participation shrinks to zero as the level of rationality of players grows.*

4 THE SYMMETRIC SETTING

We have shown that under the level-k model of reasoning, any instantiation of the Kilgour-Gerchak mechanism for which participation is unbiased yields truthful reports. We have also explored the criteria for participation in such mechanisms. A natural question is how to design mechanisms with unbiased participation. This question does not have a clean analytical answer in the general case. To gain some intuition about this question, we therefore turn our attention to one particular information structure that permits tractable analysis under the level-k model.

We consider a scenario in which signals are binary, that is, $\Sigma_i \in \{0, 1\}$ for all i, and are drawn independently for each player, conditional on the state of the world X. Furthermore, each player i's signal is "correct" with some fixed and known probability c_i, that is, $\Pr[\Sigma_i = x \mid X = x] = c_i$ for $x \in \{0, 1\}$. Finally, to simplify analysis and presentation, we assume that, a priori, $\Pr[X = 1] = \Pr[X = 0] = 1/2$. One way of viewing this assumption is that prior public information does not favor either outcome, but rather all information in favor of some outcome is received privately by the players through their signals. We refer to this setting as the *symmetric setting*.

In the symmetric setting, the posterior in (1) is simply

$$p_i(\sigma_i) = \begin{cases} c_i & \text{if } \sigma_i = 1, \\ 1 - c_i & \text{if } \sigma_i = 0. \end{cases} \tag{3}$$

4.1 FULLY CHARACTERIZING MECHANISMS WITH UNBIASED PARTICIPATION

The next two results provide matching sufficient and necessary conditions for achieving unbiased participation in the symmetric setting. First, Theorem 4 shows that to achieve unbiased participation it is sufficient to use the Kilgour-Gerchak mechanism with a scoring rule that is symmetric in the sense that $S(p, x) = S(1 - p, 1 - x)$ for all p and x, or equivalently, has associated entropy function G with $G(p) = G(1 - p)$ for all p. In this case, a player i chooses to participate at level k if and only if his expected score is higher than the average expected score of all other players who would participate at level $k-1$. Theorem 5 then shows that this symmetry is also necessary in order to achieve unbiased participation in this setting.

Theorem 4 (Sufficient condition for unbiasedness) *In the symmetric setting, the Kilgour-Gerchak mechanism with strictly proper scoring rule S exhibits unbiased participation at all levels if $S(p, x) = S(1 - p, 1 - x)$ for all $p \in [0, 1]$ and $x \in \{0, 1\}$. Moreover, player i at level k participates if and only if*

$$G(c_i) > \frac{1}{|\mathcal{P}_{-i}^{(k-1)}|} \sum_{j \in \mathcal{P}_{-i}^{(k-1)}} G(c_j), \tag{4}$$

where G is the entropy function associated with S.

Proof: The proof is by induction. By definition, all players participate at level 0, so participation is unbiased. Consider any $k > 0$ and suppose that for all levels $k' < k$, participation is unbiased. Then by the entropy characterization of scoring rules in Theorem 1 and by Lemma 1, we have that player i with signal $\Sigma_i = \sigma_i$ participates only if

$$G(p_i(\sigma_i)) > \frac{1}{|\mathcal{P}_{-i}^{(k-1)}|} \sum_{j \in \mathcal{P}_{-i}^{(k-1)}} \mathbb{E}[S(p_j(\Sigma_j), X)|\Sigma_i = \sigma_i].$$

The symmetry of S implies that $G(p) = G(1-p)$ for all $p \in [0,1]$. Using this symmetry and the fact that $p_i(\sigma_i)$ is either c_i or $1 - c_i$, we have $G(p_i(\sigma_i)) = G(c_i)$ regardless of the signal realization σ_i. To complete the proof, we then need only to show that for any $j \neq i$ and any $\sigma_i \in \{0,1\}$,

$$\mathbb{E}[S(p_j(\Sigma_j), X)|\Sigma_i = \sigma_i] = G(c_j). \tag{5}$$

Since $G(c_j)$ does not depend on player i's signal, this would imply that participation is unbiased at level k.

By exploiting symmetry as described below, we have that

$$\begin{aligned}
&\mathbb{E}[S(p_j(\Sigma_j), X)|\Sigma_i = 1] \\
&\quad = \mathbb{E}[S(1 - p_j(\Sigma_j), 1 - X)|\Sigma_i = 1] \\
&\quad = \mathbb{E}[S(p_j(1 - \Sigma_j), 1 - X)|\Sigma_i = 1] \\
&\quad = \mathbb{E}[S(p_j(\Sigma_j), X)|\Sigma_i = 0].
\end{aligned}$$

The first equality follows from the symmetry of S. The second follows from (3), which implies that $1 - p_j(\Sigma_j) = p_j(1 - \Sigma_j)$. The third can be easily verified by expanding out the expressions and exploiting the symmetry in both the prior and the signal error. Hence, we have

$$\begin{aligned}
&\mathbb{E}[S(p_j(\Sigma_j), X)] \\
&\quad = \Pr[\Sigma_i = 1]\mathbb{E}[S(p_j(\Sigma_j), X)|\Sigma_i = 1] \\
&\quad\quad + \Pr[\Sigma_i = 0]\mathbb{E}[S(p_j(\Sigma_j), X)|\Sigma_i = 0] \\
&\quad = \mathbb{E}[S(p_j(\Sigma_j), X)|\Sigma_i = 1] \\
&\quad = \mathbb{E}[S(p_j(\Sigma_j), X)|\Sigma_i = 0].
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
&\mathbb{E}[S(p_j(\Sigma_j), X)] \\
&\quad = \Pr[\Sigma_j = 1]\mathbb{E}[S(p_j(\Sigma_j), X)|\Sigma_j = 1] \\
&\quad\quad + \Pr[\Sigma_j = 0]\mathbb{E}[S(p_j(\Sigma_j), X)|\Sigma_j = 0] \\
&\quad = \frac{1}{2}G(c_j) + \frac{1}{2}G(1 - c_j) = G(c_j).
\end{aligned}$$

Combining the previous two equalities gives us Equation (5), completing the proof. ■

The next result provides a matching necessary condition. Note that for any symmetric scoring rule S, adding a constant payment that depends on the outcome x but not on the report p would break symmetry but would not affect the resulting Kilgour-Gerchak payments; since this constant amount is added to *all* players' scores, it cancels out when comparing player i's score to the average score of other participants. The necessary condition states that unbiased participation is achievable only if the scoring rule used is "equivalent to" a symmetric scoring rule in this way.

Theorem 5 (Necessary condition for unbiasedness) *In the symmetric setting, if the Kilgour-Gerchak mechanism exhibits unbiased participation for level-k players with all possible signal accuracies for all levels k, then the payments are equivalent to those using Kilgour-Gerchak with a scoring rule S that satisfies $S(p,x) = S(1-p, 1-x)$ for all $p \in [0,1]$ and $x \in \{0,1\}$.*

Proof: Assume that the Kilgour-Gerchak mechanism using scoring rule $\bar{S}$ exhibits unbiased participation for level-k players for all possible vectors of signal accuracies and all levels k.

First note that for any scoring rule $\bar{S}$ bounded in $[0,1]$, the scoring rule S defined by $S(q,1) = \bar{S}(q,1) + (1 - \bar{S}(1,1))$ and $S(q,0) = \bar{S}(q,0) + (1 - \bar{S}(0,0))$ remains bounded in $[0,1]$ and results in identical payments to each player when used in the Kilgour-Gerchak mechanism. For the purposes of this proof, we therefore consider the alternative representation of the mechanism as the Kilgour-Gerchak mechanism with this modified scoring rule S. Note that $S(1,1) = S(0,0) = 1$ and therefore $G(1) = G(0) = 1$.

Suppose there are only two players, i and j, with equal accuracies $c_i = c_j = c$, and consider the level-1 behavior of agent i. At level 0, agent j always participates. By Lemma 1, Theorem 1, and (3), we have that player i with signal $\Sigma_i = 1$ participates if and only if

$$G(c) > \mathbb{E}[S(p_j(\Sigma_j), X)|\Sigma_i = 1].$$

Expanding out the terms on both sides and rearranging, this inequality is equivalent to

$$S(c,1) + S(c,0) > S(1-c,1) + S(1-c,0). \tag{6}$$

Similarly, player i participates with signal $\Sigma_i = 0$ if and only if

$$G(1-c) > \mathbb{E}[S(p_j(\Sigma_j), X)|\Sigma_i = 0].$$

Expanding out terms in a similar way and rearranging, this condition reduces to

$$S(1-c,1) + S(1-c,0) > S(c,1) + S(c,0). \tag{7}$$

For participation to be unbiased it must be that (6) and (7) either both hold or both do not hold. Clearly they cannot both hold simultaneously. The only way that both could simultaneously not hold is if

$$S(c,1) + S(c,0) = S(1-c,1) + S(1-c,0).$$

By the characterization of scoring rules in Theorem 1, the corresponding entropy function must then satisfy

$$
\begin{aligned}
&G(c) + G'(c)(1-c) + G(c) + G'(c)(-c) \\
&= G(1-c) + G'(1-c)c + G(1-c) + G'(1-c)(c-1)
\end{aligned}
$$

which reduces to

$$G(1-c) + (c-1/2)G'(1-c) = G(c) + (1/2-c)G'(c). \quad (8)$$

Since c was chosen arbitrarily, in order for participation to be unbiased for *all* vectors of signal accuracies, this equality must hold for all $c \in [1/2, 1]$.

It remains to show that this implies symmetry in S. It must be the case that $G(c) = G(1-c)$ at $c = 1/2$ and $c = 1$. (The latter is true because we have replaced $\bar{S}$ with S.) Suppose that there is some $c \in (1/2, 1)$ such that $G(c) > G(1-c)$. Then by continuity of G, there must be some c_1 and c_2 such that

(a) $G(c_1) = G(1-c_1)$,
(b) $G(c_2) = G(1-c_2)$,
(c) $G(c) > G(1-c)$ for all $c \in (c_1, c_2)$.

Condition (c) and (8) imply that for all $c \in (c_1, c_2)$, $G'(c) > -G'(1-c)$. But this contradicts conditions (a) and (b). Therefore, there cannot exist any c with $G(c) > G(1-c)$.

A symmetric argument can be made for the case in which $G(c) < G(1-c)$ for some $c \in (1/2, 1)$, so we must have $G(c) = G(1-c)$ for all c. The characterization in Theorem 1 can be used to easily show that this implies the desired symmetry in S. ■

4.2 MAXIMIZING PARTICIPATION RATES

Unbiased participation is desirable for information aggregation. However, as shown in Corollary 1, it necessarily leads to participation shrinking to zero as players' level of rationality grows. Our final result for the symmetric setting shows how to select the scoring rule that maximizes participation among those that lead to unbiased participation. This is accomplished in the limit as the entropy function G becomes very close to linear, that is, as the scoring rule S becomes very close to being only weakly proper.

Theorem 6 (Maximal participation) *Among all symmetric strictly proper scoring rules, maximal participation can be achieved as the limit of a sequence of strictly proper scoring rules with corresponding entropy functions of the form $G(p) = |2p-1|^\beta$ with $\beta > 1$ as $\beta \to 1$.*

Proof: Consider any symmetric scoring rule with symmetric entropy function G. By Equation (4) in Theorem 4 and the convexity of the entropy function, player i at level k would never participate unless

$$G(c_i) \geq G\left(\frac{1}{|\mathcal{P}_{-i}^{(k-1)}|}\sum_{j \in \mathcal{P}_{-i}^{(k-1)}} c_j\right).$$

Since we have assumed G is symmetric around $1/2$ and strictly convex, any such function G must be increasing on $[1/2, 1]$. Since $c_j \in [1/2, 1]$ for all j, we have that at level k, player i would never participate unless

$$c_i > \hat{c}_{-i}^{(k-1)} \equiv \frac{1}{|\mathcal{P}_{-i}^{(k-1)}|}\sum_{j \in \mathcal{P}_{-i}^{(k-1)}} c_j. \quad (9)$$

This is an "only if" condition that holds for any symmetric entropy function G. To maximize participation, we would like to select G to have a matching "if" condition that is as close to this as possible.

Consider the family of scoring rules defined by entropy function $G^\beta(p) = |2p-1|^\beta$ with $\beta > 1$. By Equation (4), player i would participate at level k under this scoring rule if and only if

$$
\begin{aligned}
2c_i - 1 &> \frac{1}{\left|\mathcal{P}_{-i}^{(k-1)}(\beta)\right|^{1/\beta}} \left(\sum_{j \in \mathcal{P}_{-i}^{(k-1)}(\beta)} (2c_j - 1)^\beta\right)^{1/\beta} \\
&= \frac{1}{\left|\mathcal{P}_{-i}^{(k-1)}(\beta)\right|^{1/\beta}} \left\|\{2c_j - 1\}_{j \in \mathcal{P}_{-i}^{(k-1)}(\beta)}\right\|_\beta
\end{aligned}
$$

where $\|\cdot\|_\beta$ denotes the L_β norm and we use the notation $\mathcal{P}_{-i}^{(k-1)}(\beta)$ to emphasize the dependence of the set of participating players $\mathcal{P}_{-i}^{(k-1)}$ on β.

We first use this to show that for any $\beta, \beta' > 1$ with $\beta' < \beta$, participation declines more gradually under the scoring rule defined by $G^{\beta'}$ than it does under the scoring rule defined by G^β. We do so by induction. Since all players participate at level 0, for any i we have $\mathcal{P}_{-i}^{(0)}(\beta) = \mathcal{P}_{-i}^{(0)}(\beta')$. Assume that for all $k' < k$, $\mathcal{P}_{-i}^{(k')}(\beta) \subseteq \mathcal{P}_{-i}^{(k')}(\beta')$. By standard properties of norms, we then have

$$
\begin{aligned}
&\frac{1}{\left|\mathcal{P}_{-i}^{(k-1)}(\beta)\right|^{1/\beta}} \left\|\{2c_j - 1\}_{j \in \mathcal{P}_{-i}^{(k-1)}(\beta)}\right\|_\beta \\
&\geq \frac{1}{\left|\mathcal{P}_{-i}^{(k-1)}(\beta)\right|^{1/\beta'}} \left\|\{2c_j - 1\}_{j \in \mathcal{P}_{-i}^{(k-1)}(\beta)}\right\|_{\beta'} \\
&\geq \frac{1}{\left|\mathcal{P}_{-i}^{(k-1)}(\beta')\right|^{1/\beta'}} \left\|\{2c_j - 1\}_{j \in \mathcal{P}_{-i}^{(k-1)}(\beta')}\right\|_{\beta'}.
\end{aligned}
$$

The last line follows from the fact that since the players who choose to participate are always those with the highest quality (corresponding to higher values of $2c_j - 1$), if a larger group participates they are lower quality on average.

This implies that if a player participates at level k under the scoring rule defined by G^{β} then he also participates for any $\beta' < \beta$, completing the inductive step.

In the limit as $\beta \to 1$, we get that the participation constraint at each level k converges to the constraint $c_i > \hat{c}_{-i}^{(k-1)}$, which, from Equation 9, is the participation limit for any symmetric strictly proper scoring rule. ■

This result shows that participation levels rise as the scoring rule used becomes "closer to" weakly proper, providing experts less incentive to report their true beliefs rather than make a false report. In fact, it is easy to see that maximum participation could be achieved using the weakly proper scoring rule with entropy function $G(p) = |2p-1|$, though this would result in a loss of strict truthfulness. This result is somewhat intuitive. Since the Kilgour-Gerchak mechanism is a zero-sum game, bigger rewards for the most accurate players require bigger punishments for the least accurate, causing those with less information to drop out at lower levels of rationality. By rewarding all players more evenly, less accurate players do not drop out as quickly. However, this may have consequences in real-world scenarios in which an expert may not find it worth his time to participate if the rewards for highly accurate information are low, even if he stands to make a profit on expectation.

4.3 UNCERTAINTY ABOUT OPPONENTS

One potential objection to our model is the assumption that each player i knows the accuracy parameter c_j of every other player j. This assumption is certainly unrealistic in settings in which the number of players is large or the pool of players anonymous. We briefly remark that most of our results can be extended to the Bayesian setting in which it is necessary only for each player i to know a *distribution* over the types (in this case, accuracy parameters) of other players. However, this extension requires modifying the definition of unbiasedness so that a player's decision to participate may depend on parts of his private information (in particular, his type), but remains independent of the outcome X. To preserve the clarity of our analysis, we omit the details and present only the more simple setting here.

5 DISCUSSION

In order to design wagering mechanisms to elicit honest, unbiased beliefs from groups of experts, it is necessary to understand how experts behave. Previous analyses have assumed extreme behavior; either experts are fully rational, in which case standard no-trade theorems apply, or experts have immutable beliefs and are essentially oblivious to the fact that they are participating in a zero-sum game. In this paper, we search for middle ground, analyzing the behavior of boundedly rational level-k players who recognize they are in a game but are still overconfident in their reasoning. We examine the design implications of this model, seeking instantiations of Kilgour and Gerchak's shared scoring rule wagering mechanism that encourage unbiased and truthful participation at high rates.

This paper can be viewed as a contribution to the new but growing research area of *behavioral mechanism design* (Ghosh and Kleinberg, 2014; Easley and Ghosh, 2015) in which insights from behavioral game theory are applied to design mechanisms tailored to real (or at least more realistic) human participants as opposed to idealized rational agents. Of course any theory is only as good as the model on which it is based. While the behavioral game theory literature provides support that the level-k model is a decent predictor of human behavior in game theoretic settings—including one-shot betting games in which the no-trade theorem would typically apply (Brocas et al., 2014)—additional experimental work is needed to understand how well it models the behavior of real experts participating in wagering mechanisms like Kilgour-Gerchak. Still, we believe that our analysis takes a valuable first step towards understanding the ability of wagering mechanisms to aggregate information from experts who are neither fully rational nor fully naive.

References

R. J. Aumann. Agreeing to disagree. *The Annals of Statistics*, 4(6):12360–1239, 1976.

A. Bosch-Domènech, J. G. Montalvo, R. Nagel, and A. Satorra. One, two, (three), infinity, ...: Newspaper and lab beauty-contest experiments. *American Economic Review*, 92(5):1687–1701, 2002.

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

I. Brocas, J. D. Carrillo, S. W. Wang, and C. F. Camerer. Imperfect choice or imperfect attention? Understanding strategic thinking in private information games. *Review of Economic Studies*, 81(3):944–970, 2014.

C. F. Camerer, T.-H. Ho, and J.-K. Chong. A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119(3):861–898, 2004.

S. Chun and R. Shachter. Strictly proper mechanisms with cooperating players. In *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2011.

M. Costa-Gomes, V. P. Crawford, and B. Broseta. Cognition and behavior in normal-form games: An experimental study. *Econometrica*, 69:1193–1235, 2001.

M. A. Costa-Gomes and V. P. Crawford. Cognition and behavior in two-person guessing games: An experimental study. *American Economic Review*, 96(5):1737–1768, 2006.

V. P. Crawford and N. Iriberri. Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions? *Econometrica*, 75:1721–1770, 2007.

P. Dubey, J. Geanakoplos, and M. Shubik. The revelation of information in strategic market games: A critique of rational expectations equilibrium. *Journal of Mathematical Economics*, 16:3105–137, 1987.

D. Easley and A. Ghosh. Behavioral mechanism design: Optimal crowdsourcing contracts and prospect theory. In *Proceedings of the 16th ACM Conference on Economics and Computation (EC)*, 2015.

J. D. Geanakoplos and H. M. Polemarchakis. We can't disagree forever. *Journal of Economic Theory*, 28(1):192–200, 1982.

A. Ghosh and R. Kleinberg. Behavioral mechanism design: Optimal contests for simple agents. In *Proceedings of the 15th ACM Conference on Economics and Computation (EC)*, 2014.

T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

T.-H. Ho, C. F. Camerer, and K. Weigelt. Iterated dominance and iterated best response in experimental "p-beauty contests". *American Economic Review*, 88(4): 947–69, 1998.

J. Jordan and R. Radner. The nonexistence of rational expectations equilibrium: A robust example. Working paper, Department of Economics, University of Minnesota, 1979.

D. M. Kilgour and Y. Gerchak. Elicitation of probabilities using competitive scoring rules. *Decision Analysis*, 1(2): 108–113, 2004.

N. S. Lambert, J. Langford, J. Wortman, Y. Chen, D. Reeves, Y. Shoham, and D. M. Pennock. Self-financed wagering mechanisms for forecasting. In *Proceedings of the 9th ACM Conference on Electronic Commerce (EC)*, 2008.

R. E. Lucas. Expectations and the neutrality of money. *Journal of Economic Theory*, 28:103–124, 1972.

R. D. McKelvey and T. R. Palfrey. Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38, 1995.

P. Milgrom and N. L. Stokey. Information, trade and common knowledge. *Journal of Economic Theory*, 26(1): 17–27, 1982.

R. Nagel. Unraveling in guessing games: An experimental study. *American Economic Review*, 85(5):1313–1326, 1995.

L. T. Nielsen, A. Brandenburger, J. Geanakoplos, R. McKelvey, and T. Page. Common knowledge of an aggregate of expectations. *Econometrica*, 58(5):1235–1238, 1990.

C. R. Plott and S. Sunder. Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometrica*, 56:1085–1118, 1988.

B. W. Rogers, T. R. Palfrey, and C. F. Camerer. Heterogeneous quantal response equilibrium and cognitive hierarchies. *Journal of Economic Theory*, 144(4):1440–1467, 2009.

L. J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971.

D. Shapiro, X. Shi, and A. Zillante. Level-k reasoning in a generalized beauty contest. *Games and Economic Behavior*, 86:308–329, 2014.

Y. Sløvik. Strength of dominance and depths of reasoning – An experimental study. *Journal of Economic Behavior and Organization*, 70(1–2):196–205, 2009.

D. Sonsino, I. Erev, and S. Gilat. On rationality, learning and zero-sum betting – An experimental study of the no-betting conjecture. Technion working paper, 2001.

D. O. Stahl and P. W. Wilson. Experimental evidence on players' models of other players. *Journal of Economic Behavior and Organization*, 25(3):309–327, 1994.

D. O. Stahl and P. W. Wilson. On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1):218–254, 1995.

J. R. Wright and K. Leyton-Brown. Beyond equilibrium: Predicting human behavior in normal form games. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI)*, 2010.